MAPPING AND EVALUATION OF MACHINE PERFORMANCE DATA FROM FARM OPERATIONS

BY

THOMAS LEE KENNEDY

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Agricultural and Biological Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Adviser:

Professor Alan Hansen

ABSTRACT

The costs associated with inefficient agricultural operations are increasing as the price of fuel, labor, and machinery increases. This study presents an analysis of the machine performance data collection method performed on two large-scale Midwestern farms. This data collection method sought to record data from each machine involved in agricultural production on a one second interval for one season. After the collection of the data, it was analyzed on a data quality basis and a machine performance basis. The objective of data quality analysis was to create a metric on which the confidence of the results of the machine performance analysis could be based. In analyzing machine performance, the geographic information system, ArcGIS™, was used to present the information in a geographic reference.

Results of the analysis of the data collection process revealed several flaws that reduced the quality of the collected data. The cause of the data quality issues associated with the collection of the data was attributed to the transmission of the data via cellular networks instead of an error with the actual recording of data from the machine. Additional recommendations regarding the setup of data loggers were made in order to increase the detail of the data, and increase the value of conclusions that could be made based on these data.

After reviewing the process with which the data were collected, a method to analyze the data was created. This method included a visual inspection of machine parameters for geographic trends in the data. Additionally, the data was analyzed statistically to determine the relationships between machine parameters on a single machine. The visual analysis of parameters on multiple machines was able to show the interaction of independent machines in field operations. A comparison of machine performance was provided on an individual field basis and a farm wide

basis in order to identify fields in which machine performance differed significantly from the farm average. Combining the method included in this study with improved data quality resulting from the implementation of data collection recommendations was anticipated to provide information to improve machine design and to direct farm managers to areas of inefficiencies in the farming operation and the operation of individual machines in the fields.

ACKNOWLEDGEMENTS

This research was financially supported by the John Deere Technology Innovation Center. The data collection method and the data collection itself were designed and performed by John Deere as part of a larger study of agricultural operations efficiency, and without that data, this study would not have been possible. Dr. Alan Hansen provided an unquantifiable amount of assistance and guidance in the completion of this project. The additional support of Dr. Richard Gates and Dr. Prasanta Kalita in guiding the writing process of this document was of great value. I thank these professors for their time and effort assisting me in my academic endeavors.

The support of my family throughout my academic career has been an invaluable resource. I cannot thank my grandfather, Dale Spore, enough for providing me with the inspiration to become an engineer and for teaching me the thought-based approach to solving problems: “measure twice, cut once.”

Although my family provided inspiration and support for my academic endeavors, I cannot neglect the support of my fiancée, Jessica Lipski. She, above all, assisted me in persevering during difficult times in school to reach the finish line.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Increases in fuel prices, cost of labor, and other production costs related to agriculture have increased the demand for solutions to increase the efficiency of agricultural operations. Research into the operation of agricultural equipment has shown that on average farmers use only 60% of their tractor's rated engine capacity and that machine operators could save between 11.3% and 20% of fuel costs by shifting to different transmission gears to operate their equipment more efficiently (Grogan et al. 1987). The addition of the United Sates Environmental Protection Agency's Tier 4 engine emissions standards to already strict engine emissions regulations has provided additional need to address agricultural machine inefficiencies.

In order to compensate for the inefficient costs associated with improper machine sizing, operation, and logistics in agricultural production, John Deere is undergoing a project focusing on improving the efficiency of agricultural operations by identifying and recording fuel usage and other operational parameters. The research included in this document is a small portion of the larger study being performed by John Deere. This project attempts to provide recommendations in accordance to the overall objective of the John Deere study with a focus on analyzing one year's worth of data regarding production agriculture on two Midwestern farms. Specifically, it focuses on the evaluation of the data collection procedure and the creation of a methodology to interpret the data that were collected by John Deere.

CHAPTER 2: OBJECTIVES

Agricultural operations are both economic and labor intensive processes. In order to provide their customers with a high quality product, John Deere has collected information regarding the agricultural operations of two independent Midwestern farms. The objective of this study was to evaluate the data collection method used by John Deere to record information about one year's worth of agricultural operations in conjunction with developing a method, through which, the information could be interpreted.

The data collection method performed by John Deere was analyzed in comparison to literature published on analyzing large and complex data sets as well as literature published on analyzing data sets for quality and completeness. This portion of the study was completed in order to make recommendations to John Deere to improve the collection and management of information collected by John Deere.

The information that was collected by John Deere was used to create a structure through which new data from subsequent years of collection could be analyzed from season to season. The specific objectives of the data analysis process were to:

- Identify anomalies in machine performance
- Determine relationships between machine parameters
- Provide explanations of those connections to both John Deere and the farm managers from which the data were collected
- Develop both design and operational recommendations for improving agricultural operations based on the data
- Compare operations in terms of performance metrics

CHAPTER 3: LITERATURE REVIEW

Recent increases in fuel prices, cost of labor, and regulation on engine emissions have increased the importance of research into increasing the efficiency of agricultural equipment. This study evaluated the performance of in-field agricultural machinery as part of a larger study by John Deere that focused on a broad range of possible increases in agricultural efficiency. In order to evaluate the data collection process, it was determined that two major subjects needed to be addressed. First, the collection of large amounts of data required that procedures be researched and subsequently established for the data collection and management processes. Strategies for managing and processing large data sets were taken from agricultural disciplines when possible, but research from many domains outside of agriculture were analyzed due to the multidisciplinary nature of many problems with managing and processing large data sets. Second, a method of determining data quality was implemented to prove that decisions made through the analysis of these data were made based on an accurate representation of the machines' performance. The evaluation of data quality was determined to be very dependent upon the type of data collected and the process through which the data were interpreted. In order to create a custom method of data quality evaluation, previous studies on agricultural machine performance were reviewed. In order to improve upon these methods, the strategies for evaluating data quality were also evaluated from fields unrelated to agriculture but that have experienced similar data quality problems.

After completing research into managing and processing large data sets, a strategy to analyze the machine performance data was developed based on research into the two additional components of the study. The capabilities of ArcGIS™, a program used to analyze geospatial information, were evaluated for possible use in machine analysis. To complement the research

into capabilities of ArcGIS™, published studies on machine performance analysis were reviewed in order to determine procedures that both should, and should not, be duplicated in this study. The research evaluated according to these four subjects is presented in sections 3.1, 3.2, 3.3, and 3.4 according to their general subject: "Complex Data Sets," "Data Quality," "ArcGIS™ Capabilities," and "Machine Performance," respectively.

3.1 Complex Data Sets

Studying the in-field performance of agricultural machinery was a complex task that was further complicated by the size of the data sets that were recorded. Jacobs (2009) discovered that some strategies for interpreting relatively small data sets have operated very efficiently and effectively, but they often quickly became overwhelmed when data included temporal and spatial information that had to be processed simultaneously. According to Hao et al. (2008), the first priority in creating an overall strategy for dealing with data complexity should be determining the method through which the data were to be displayed and analyzed. After determining the desired use of the data, the various components of complexity of that data should be identified in order to create a system that addressed those issues of complexity (Rzevski 2011). The characteristics of the data were evaluated to determine the time and effort required of each process to analyze the data in an effective and timely manner. The most time and resource intensive steps in the analysis, referred to as "bottlenecks," were identified in order to develop strategies to cope with these limitations (Yan et al. 2011). After identifying bottlenecks in the process, the structure of the process could be altered in a way that took advantage of any combination of user abilities, software characteristics, and computer speed (Hawick et al. 2003).

3.1.1 Data Analysis Processes

The first step to create a system to analyze large amounts of data should be the creation of a generic description of who would access the system, how they would access the system, what type of knowledge they may possess, how they may manipulate the system, and what their objectives are (Kimmance et al. 1999). In addition to designing a data analysis process based upon the user, the type of data that is to be analyzed is a very crucial constraint on the type of structure (Hao et al. 2008).

An evaluation was provided regarding the use of density displays for the analysis of two different types of data that were both temporal in nature. The first data set related to the loads placed on individual central processing units (CPU) in a server bank. For the operator of a server bank, it was important to identify individual CPUs that operated either close to full capacity or at a relatively low capacity in order to better manage the allocation of load. The second set of data contained sales Figures for a store. The data were depicted using two forms of density display that showed the temporal progression of data. The display utilizing the shifting method displayed information by placing the most current data in the column to the far right. Then, it shifted the data one column to the left to make room for new information that will replace it in the farthest right column upon the generation of a newer data set. The second method that was used in displaying the data used a circular overlap method. In this display type, the first set of data collected was placed in the column furthest to the left. As new information was collected, it was displayed in the next column to the right. When the number of columns in the display was completely filled, the next information to be collected was rewritten over the oldest information being displayed (Hao et al. 2008).

Following the creation of these displays, a group of users was surveyed from various sections within Hewlett Packard Laboratories regarding the preference of each method with regards to finding patterns and finding anomalies in the data. The shifting method was given a score of 2.5 out of 3, with three being the best rating, for finding patterns and a 1.25 out of 3 for finding anomalies. The circular overlay method was given a 2.35 and 2.625 out of 3 for finding patterns and finding anomalies respectively (Hao et al. 2008). This survey proved the initial indication that the characteristics and objectives of the user tend to influence the type of structure in which the data was presented (Kimmance et al. 1999).

Although building a structure for interpreting data in the case of this study could have been simplified due to the fact that the structure was being designed around two separate farms, the structure was designed to be able to fit stakeholders in an operation that may have completely different objectives. Instead of identifying a structure that allows one particular group to achieve their objectives, it was better to create a structure that allows for the change of objectives with different users. Three steps were proposed in creating model structures for large data sets that may have different objectives: building, analyzing, and maintaining. To accommodate various user profiles and objectives, the automated or semi-automated creation of models within a given structure should allow for constraints to be applied that are specific to that model (Liu and Tuzhilin 2008). The use of ArcGIS™ to automatically generate a set of maps that are common amongst all models with the ability to add or subtract constraints allowed the user to view the output of the model with their specific objectives prioritized. The ability of the user to observe patterns and anomalies specific to their objectives in a visual manner helped the user avoid overlooking portions of the data that may not fit within specific structure constraints that worked with another data set and user (Pundt and Brinkkötter-Runde 2000). The analysis stage for

designing model structures suggested the creation of a feature to query models that have previously been created and used based on the generic structure as well as analyzing those models for effectiveness in addressing user objectives. In conjunction with the creation of querying features for the analyzing stage of model design, the maintenance of those models for future use was also a priority. The comparison of current models to past models and the addition of new information to previous models assisted in model verification and validation (Liu and Tuzhilin 2008).

3.1.2 Determining Data Complexity

The degree to which information is or is not complex is not only dependent on the actual information contained within a data set. It is also heavily dependent on a combination of how the data is used and how individual pieces of data are related to each other. Rzevski (2011) hypothesized that a data set and system could be considered to be complex if the majority of the following seven characteristics were present:

- Interdependency – The agents within a system are not connected through dependent relationships.
- Autonomy – There is no central controlling authority for the agents within a system, but the agents do follow a general set of rules.
- Emergence – The behavior of a system as a whole is dictated by the sum of the actions of agents within that system.
- Far From Equilibrium – The diverse inputs from the agents within the system change with a frequency that never permits the system to reach an equilibrium state.

- Nonlinearity – The relationship between the agents can have varied effects on the system as a whole due to the possibility that agents' actions can have a compounding effect in increasing the amplitude of each agent's influence on the system.
- Self-Organization – Although the system is far from equilibrium, the system reacts to perceived need within that system.
- Co-Evolution – The system evolves with its environment in an irreversible manner.

Although these characteristics of complexity are descriptive of a system, they are also descriptive of the data generated by a system that can be used for creating system models and analysis tools. A system's complexity was discussed in terms of the global market. Like the farms being analyzed in this study, the global market is made up of independent agents that influence the state of the system. To create an accurate model of the global market, the suppliers, consumers, manufacturers, and other independent agents could not be simplified to have a specified overall effect on the system (Rzevski 2011). Similarly, weather, fuel prices, and breakdowns could not be simplified to have only one effect on farming.

In order to preserve the integrity of the information gathered from the complex agricultural system within this study, the information was to be processed within the system model such that information was not over-simplified. The manner in which the data were processed with the computer had an immense impact on the characterization of the system. An artificial data set representing the information that could be gathered in a world-wide population census was created by Jacobs (2009) with the goal of determining the median age by gender for each country. This data set included several pieces of information including a 7 bit field for age with 128 possible values, a 1 bit field for gender, and an 8 bit field for 256 countries (approximately 192 member states of the UN). After writing a basic code with "bins" for each of

the possible combinations of age, gender, and country (65,536 in total), the processor for the computer simply read the data one by one and assigned a newer value to each bin. Jacobs was able to accomplish the objective in approximately 15 minutes with a consumer grade desktop computer. Jacobs (2009) then repeated the same calculation on "a commonly used enterprise-grade database system (PortgreSQL) running on […] an eight-core Mac Pro workstation with 20GB RAM and two terabytes of RAID 0 disk." Despite the extreme advantage of the second system in computing power, the second program artificially inflated the actual size of the data set by storing the information as 32 bits and aborted the process after six hours (Jacobs 2009).

Similar complications arose with the use of ArcGIS™ for data processing. Common types of information storage in ArcGIS™ include vector, raster, and point data. In operations performed on raster data, processor load was much less than for those processes that were performed on vector or point data. Given the ability to convert vector or point data to raster data, converting to raster data implies that the speed and efficiency of processing the data would increase. However, the inability to convert back to point or vector data from raster data as easily as converting point and vector data to raster data can cause problems when creating model structures for the analysis of agricultural operations (Hawick et al. 2003). Parameters that describe agricultural operations are not always based upon numeric values. Often, the description of agricultural operations includes verbal descriptions. The record of these characteristics may be more complicated to define on a common scale. Analyzing the data in ArcGIS™ required the system architect to choose between a data format that can be processed on a more simple system or point and vector data that may require the use of parallel processing applications (Hawick et al. 2003; Jasiewicz 2011).

3.1.3 Data Processing Bottlenecks

In any process that evaluates information either sequentially or in parallel, there is typically one step that limits the speed at which the entire process can operate. This step, referred to as a bottleneck, is often the most resources intensive or time consuming portion of data processing (Jasiewicz 2011). If a model structure is being optimized for the quickest performance, the overall operation can only improve in speed if the performance of the slowest portion of the structure is improved. In the example of data processing using a consumer grade desktop computer to calculate the median age by gender for each country, the overall operation time was governed by the speed at which the data could be read from the disk (90MB/s). This process was such a significant bottleneck in the overall process that the structure was, "shamefully underutilizing the CPU the whole time" (Jacobs 2009).

When a functional diagram is easy to create and time for each function is easy to determine, the identification of the bottleneck can be relatively simple compared to the creation of a solution to the bottleneck. When viewing data sets that contained gigabytes to terabytes worth of information in ArcGIS™, the process of drawing all of the data within that set took hours to complete when the data was viewed at a full extent (Alkobaisi et al. 2012). Despite the statement that information should not be generalized when processing that information through a data processing structure (Rzevski 2011), the generalization of data at certain scales was determined to be an effective tool in minimizing the effects of drawing maps at different views. If individual data points were preserved in the model for individual inspection at close scales, the generalization of data points at large scales did not remove any ability from the user to interpret that data. The user did not gain any additional information from a state-wide scaled map if a cluster of buildings were represented as individual building data points or if they were

represented as a city. In addition to cities, this method of generalizing information at large extents was useful for simplifying continuous terrain features (roads, coastlines, terrain elevation lines, etc.) as a series of line segments that draw much faster than their high-quality drawings would at large scales (Chaudhry and Mackaness 2010).

Bottlenecks in a structure for the useful interpretation of large data sets were not exclusive to the processing stage. Frequently, the bottleneck occurred in either the transmission or storage stage. With modern data collection systems utilizing wireless networks to transmit information that they collect, wireless networks were typically found to be an element of the data analysis process that was easy to identify, but difficult to analyze. Wireless networks were found to be subject to a plethora of problems. Wireless networks that utilized frequencies in ranges common to other electronics were complicated due to interference between multiple signals. Depending on the structure of the network, bottlenecks were attributed to the transition from one antenna to another slowing transmission speeds. If the signal was lost for a period of time, the data transmission overloaded the available bandwidth of the system. Therefore, the wireless network to be chosen for the transmission of a large amount of data should be selected such that the wireless network could handle not only average loads for the data transmissions but also loads that would be able to return the load for the network to an equilibrium state in the event of backlogged amounts of data (Yan et al. 2011). In the first year of data collection for this study, a cellular device with minimal on-board storage was used to transmit the data to an off-site location for data storage.

3.2 Data Quality

Decisions made based upon the collection of data are only as valid as the quality of the data allows them to be (Marinos 2004). Research indicated that approximately 75% of major

companies have significant financial losses due to inaccurate billing, lost sales, and other financial errors resulting from poor data. Additionally, only one-fifth of major company executives displayed a high confidence level about their company's data (Marinos 2004). This lack of confidence in data was due to the difficulties involved in recording, transmitting, storing, and interpreting the data. These complexities involved in the use of data often led data users to ignore issues of data quality because they were unable to clearly define data quality in the context of their own use (Devillers et al. 2007). Clearly defining data quality in the context of its specified use was the first step in resolving issues related to poor quality data. Having established definitions of data quality in the context of specific uses, strategies to increase the quality of data or to cope with poor quality data can be implemented in order to increase the benefits of decisions made based upon the data (Marinos 2004; Parssian et al. 2004; Devillers et al. 2007; Kumi-Boateng and Yakubu 2010).

3.2.1 Describing Data Quality

The definition of data quality is a very vague concept due to the constraints of each user, and common standards for data quality and metrics for quantifying the quality of data do not exist (Paradice and Fuerst 1991). The use of Google Earth ™, for example, may provide adequate quality data to an individual who wants a visual representation of what a specific place looks like from above, but a user that desires detailed information regarding elevation or position may find that the quality of Google Earth ™ is not sufficient for that specific application. Kumi-Boateng and Yakubu (2010) determined that the term "quality" in the context of data is more accurately a function of the use of the data than a function of the data itself.

As each definition of data quality is unique to the specific use, generic definitions based on other users' work are often useful in creating a definition for a new data quality issue. Two

commonly used defining factors for data quality were accuracy and completeness. The first term relates to the data's ability to display a reliable and trustworthy representation of the real characteristics of the source that it is trying to describe. The second term relates to data's relevance to the user's desired outcome. A set of data could be extremely accurate, but could still be of poor quality for use in a study. If the purpose of a study was to provide the most accurate picture of a given source, then a source with a higher degree of accuracy would be preferable over a source that has a high degree of completeness and low accuracy. Conversely, if the goal of the study was to sample the highest percentage of the population, a source that is more complete would be preferable over a source that is slightly less complete but has slightly higher accuracy (Parssian et al. 2004).

In the context of geospatial data, several studies defined the issue of accuracy in more detail. Since geospatial data have more than one dimension, it was important to define accuracy in context to each category of possible accuracy. The need to discriminate between positional, temporal, and attribute accuracy was crucial in assessing the fitness for use of data (Kumi-Boateng and Yakubu 2010; Li et al. 2012). These independent categories of accuracy influenced the quality of the data in relation to the user's need. The position of a given point can be inaccurate to any particular degree, but if the user's desire was to determine how an attribute changes from point A to point B, the data could still be useful if the spatial inaccuracies were consistent among all of the data points (Kumi-Boateng and Yakubu 2010).

One final aspect of data quality that has been proposed for geospatial data is logical consistency. This descriptor of data quality described the degree to which any given piece of data conforms to a set of rules or logic that typically govern that particular type of data (Devillers et al. 2007; Kumi-Boateng and Yakubu 2010). The logical consistency of data combined with the

three components of accuracy (positional, temporal, and attribute) and completeness describe the "Famous Five" characteristics that were proposed to define the overall quality and fitness for use of geospatial data (Devillers et al. 2007).

3.2.2 Data Quality Control and Quality Assurance

Two important factors in producing a scientifically sound set of data were determined by the United States Environmental Protection Agency (2003): "Quality Control" and "Quality Assurance" (United States Environmental Protection Agency 2003). "Quality Control" referred to a procedure through which all aspects of the quality of a process can be analyzed. "Quality Assurance" included the means through which the quality of data can be assured to the end user or customer (United States Environmental Protection Agency 2003). Moody et al. (2006) documented the process of controlling and assuring quality for "… Monitoring Gaseous and Particulate Matter Emissions from Broiler Housing" (Moody et al. 2006). The documentation to ensure proper assurance of quality contained a long list of data quality standards that were determined to be acceptable. The methods in which the data were collected, how the quality of the data would be measured, who was responsible for each aspect of data quality monitoring, and an assortment of other detail regarding project management were included in this document. By documenting these processes, Moody et al. (2006) was able to provide both documentation and accountability for the standards associated with the project (Moody et al. 2006).

In order to accomplish these two aspects of quality, the Louisiana Department of Transportation and Development (2008) has published a plan of "The 5 C's" in order to describe the requirements for quality of road construction plans. These "5 C's" include being "Complete," "Consistent," "Clear," "Correct," and "Constructible." In order for a construction project to follow their quality control and quality assurance plan, any construction plans must be complete

in detail, consistent with plans for other projects, clear in their presentation of information, correct in the design, and able to be constructed (Louisiana Department of Transportation and Development 2008).

In terms of this project, there was not an object to be constructed, but "The 5 C's" as determined by the Louisiana Department of Transportation and Development (2008) still apply. Similarly, a plan for determining data quality, documentation, and accountability should be formed in order to assure the quality of the data and the subsequent quality of the results as was created by Moody et al. (2006). The plan for collecting geospatial data should be complete in detail regarding all aspects of data collection. It should be consistent with other projects in which geospatial data are collected. It should be clear in the design with no omission of detail regarding data collection. The plan should not contain any design errors that may delay the project or impede the collection of data in any other way. Though no object was to be constructed in this study, the data collection process should still be constructible or executable in the actual collection of data. The United States Environmental Protection Agency (2003) provided direction for geospatial information collection in a generic quality control plan. They suggested the clear definition of all aspects of the data prior to the execution of any data collection as part of a plan for collecting data (United States Environmental Protection Agency 2003).

3.2.3 Resolving Data Quality Issues

Resolving issues with poor quality data is not as easy as applying a rigid framework of standards to data. The uniqueness and complexity of geospatial data further prevents the user from applying a very general set of rules to the data to create a data quality metric. Though metrics cannot typically be attained easily, quality assurance is extremely important for those

creating, publishing, and releasing data sets because the end user can only make reliable decisions based on the reliability of the data (Tong et al. 2011).

When creating an accurate measure of data quality, the ability to track the history of the data though detailed documentation about each of the steps in the data organization process was determined to be extremely important. This included the process of data creation, data utilization, and the organization and maintenance of the database (Kumi-Boateng and Yakubu 2010). This was especially important in the use of geospatial information databases. This type of data can be very expensive to create, and therefore, the data required for a user's objective may not be created in the user's study. Instead, geospatial information is often created by an organization for the use of several third party organizations (Kumi-Boateng and Yakubu 2010; Tong et al. 2011; Li et al. 2012).

The ability of multiple users of geospatial information to adapt information to their own situation allows those users to make decisions that may not be based on a complete understanding of the data. Consequently, they may make inaccurate decisions based upon those misunderstandings (Devillers et al. 2007). Laskey et al. (2010) described the United States Department of Defense as a prominent example of the misuse of data in this way. The Department of Defense National Geospatial Agency publishes maps for the US military to use to make decisions on the battlefield. Two commonly used types of data include elevation and terrain feature data. The terrain features have been combined with elevation data to give battlefield commanders an estimate of the ease of movement for both their troops and enemy troops over a given terrain. Several different military algorithms interpreted the information in elevation and terrain maps to create a map that was color coded to show the commanders how quickly each portion of the terrain could be traversed. The problem with this type of map was

that the high resolution could be misleading to the end user. Although one map showed a particular point to be traversable at greater than 45 km/h, the map did not take into account the possibility that the surrounding points may only be traversable at 0-3 km/h (Laskey et al. 2010).

The quality control and assurance procedures recommended by the United States Environmental Protection Agency (2003), Moody et al. (2006), and the Louisiana Department of Transportation and Development (2008) agreed that clear, detailed planning for the project is the most important step to ensure quality. Several sources agreed that the most conservative solution to the problem of data quality issues after the well planned data collection was the use of expert opinion in conjunction with quality audits of both data sets and the processes through which they were created, processed, stored, and interpreted (Paradice and Fuerst 1991; Devillers et al. 2007; Nahm et al. 2008; Laskey et al. 2010; Tong et al. 2011; Tong and Wang 2012). Although rough sets of rules were applied through automation, frequently the rules for data quality analysis generated the same types of errors inherent to the methods with which the data was collected (Devillers et al. 2007). Automated sets of rules were not always reliable with geospatial data because it was very difficult to quantify the completeness of the data and possible mismembership in data. Expert judgment was therefore the most reliable way to ensure the quality of data (Parssian et al. 2004).

3.3 ArcGIS™ Capabilities

ArcGIS™ is a computer program designed to perform a plethora of functions pertaining to geospatial information. The many uses of ArcGIS™ include the visualization of geospatial information through maps by combining one or more layer of information. It has been used to analyze relationships between geospatial data, and it has the capability to perform many more specific functions that enable the user to gain a deeper understanding of their data and make

well-informed decisions (Gorr and Kristen 2011). The ability of ArcGIS™ to interpret different types of geospatial data is very diverse. ArcGIS™ was used to study pollution levels in the city of Prague, Czech Republic, where it was able to combine air, water, landscape, waste and noise pollution into one model. This model was then used to help both the national government in the Czech Republic and the city governments of Prague manage and reduce pollution and its related effects (Matějíček et al. 2006).

Instead of using ArcGIS™ in the analysis of geospatial environmental issues, it has also been useful in determining the relationships between two parameters that may not specifically require the use of geospatial location. The statistical tools within ArcGIS™ were used to calibrate a photo sensor on a blueberry harvester to measure yield (Chang et al. 2012). The correlation of percentage of blue pixels in a photographic monitor with yield combined environmental characteristics with machine sensor readings in order to design a new yield monitoring system (Duttmann et al. 2013). ArcGIS™ has proven to be a powerful and diverse tool to analyze geospatial information and to make well-informed decisions based on those analyses.

3.3.1 Mapping

The ability to visualize geospatial information through maps generated in ArcGIS™ gives users the opportunity to recognize qualities of data that may be missed through traditional data processing (Alkobaisi et al. 2012). ArcGIS™ allows the user to georeference their data to existing projection standards. These projection standards allow a set of data to be viewed in the same frame of reference as other layers including elevation, satellite imagery, road maps, political maps, and other types of geographic information (Clemmer 2010; Gorr and Kristen 2011; Ormsby et al. 2010). Chang et al. (2012) and Farooque et al. (2013) developed an

automated yield monitoring system for blueberry harvesters. In this study, ArcGIS™ maps were utilized to visually inspect the bare spots in blueberry fields and the corresponding yield within those fields. The maps for two different fields displayed the fruit yield and geographic locations of the bare spots within the fields. Although logic implied that fruit yield would be minimal in areas where there were no plants, the illustration of the data allowed the users to verify the assumption that bare spots corresponded to low yields visually, without an in-depth analysis of the data. Additionally, this logic verified that the photographic yield monitor was not registering a yield of blueberries in a portion of the field where there were no berries (Chang et al. 2012; Farooque et al. 2013).

This mapping technique has been effective in the visualization of variables in agriculture including weather, soil type, management practices, and land use (Resop et al. 2012). ArcGIS™ is not limited to mapping variables that are present within an input data set. Other information can be mapped that is dependent on data originally entered into ArcGIS™. This capability of ArcGIS™ was used to map the areas within a field that were traveled by machines during silage harvest. GPS measurements were taken from a position on the machine that was centered with the axle of harvesters and supporting vehicles to represent the path that the machines traveled in the field. The path of the center of the machine's axle was processed in combination with the physical dimensions of the machine's wheelbase to determine the physical location of contact between the tires of the machines and the soil (Duttmann et al. 2013).

3.3.2 Statistics and Modeling

In addition to mapping geospatial data, ArcGIS™ has tools that can be used to generate statistical characteristics of the data and tools that can be used to model parameters from the raw data. The mapping of wheel tracks was only one portion of the results that was generated from

the study of soil compaction caused by wheel tracks in the field. The information regarding machine location was then combined with the base weight of the machine, amount of fuel on board, and other machine specific parameters to determine the load applied by each wheel on the soil. The load that each wheel applied to the soil enabled the soil compaction to be calculated. With individual machines' contributions to compaction being determined, a composite map of all of the machines' paths was created. After applying statistical tools from ArcGIS™ to the data set describing maximum pressure placed upon the soil, it was determined that 62.8% of the field's area was traveled with the harvester and the supporting vehicles, and 66% of those wheel tracks were traversed more than twice (Duttmann et al. 2013).

Farooque et al. (2013) used similar statistical tools to those used in analyzing the compaction caused by machinery in harvesting silage to develop a photographic yield monitor to evaluate the relationship between plant height, elevation, slope and yield of blueberries. The statistical significance of each variable's effect on yield was determined by analyzing the geographic position of each variable. These analyses were able to associate approximately 15-35% of the variability in yield to elevation and slope alone (Farooque et al. 2013). Similar studies attempting to determine the influences of elevation, slope, and curvature surfaces have shown that up to 78% of the spatial yield variability in terraced fields could be explained through those three factors (Schmidt et al. 2003). In a similar study of the effects of the elevation, slope, aspect, and curvature on yield, three years of cotton yield and two years of corn yield showed a statistical significance between the aforementioned land characteristics and the yield (McKinion et al. 2010). These results are not difficult for statistics software to generate if the data points for yield have a corresponding value of each topographical characteristic. However, when incorporating the yield variables from multiple years, the data points collected by the harvester

regarding yield were not typically the exact same values of latitude and longitude as the previous or following years. ArcGIS™ allows data points to be interpolated between two known values to get an approximate value for any point (Matějíček et al. 2006).

When determining the suitability of land for any particular use, descriptive information of that land was not always a numeric value. Some variables had numeric values whereas others, like soil type, had a verbal classification. In order to analyze how different factors influence the overall suitability of a piece of land, Multi Criteria Decision Analysis, MCDA, was used (Chen et al. 2010). ArcGIS™ enabled the user to specify what MCDA model was desired. By assigning different factors, ratings, weights, and scores to the different input values, an overall assessment of land quality was made from a combination of both numeric and verbal data (Mendas and Delali 2012).

3.4 Machine Performance

When evaluating the performance of agricultural machinery, two separate approaches can be taken. Agricultural machinery follows established principles that describe their operating characteristics on a generic basis. Calculations of power, draft, efficiencies, etc. are easily calculated using simple equations governing those properties (Goering et al. 2006; Srivastava et al. 2006; Goering and Hansen 2008). Although this approach to evaluating machine performance can be useful in determining how a machine's setup should be changed or how the machine could be operated more efficiently, this type of analysis often narrowly focuses on the machine, neglecting outside influences. In order to evaluate machine performance, a holistic approach was suggested by several studies (Grogan et al. 1987; Yule et al. 1999; Schmidt et al. 2003; Boon et al. 2005; Yahya et al. 2009; Singh and Singh 2011). By collecting data from the field operations of agricultural machinery, the impacts of soil properties (Boon et al. 2005), operator behavior

and skill level (Grogan et al. 1987; Yahya et al. 2009), and economic factors related to agriculture (Grogan et al. 1987; Yule et al. 1999) were determined in a much broader scale than with traditional theory based research. These factors were all important parts of the study of agricultural machinery because farmers and operators may not be able to or want to optimize one portion of the operation if it sacrifices more value on the operation as a whole than it gains in value for that portion.

3.4.1 Theory Based Research

The fundamental relationships between properties of agricultural machines and their operating principles are an important starting place for the analysis of performance in field operations. The primary relationships that were of use to this study in evaluating the overall performance of agricultural machinery include, but are not limited to, those relating to power, fuel consumption, draft, and efficiency.

3.4.1.1 Calculation of Power

To determine the power output at any particular part of the machine's powertrain (flywheel, power take-off, axle, and other locations), the speed of the shaft and the torque being transmitted to the shaft can be related by (Goering and Hansen 2008)

$$P_b = \frac{2\pi T_b N_e}{60{,}000} \tag{1}$$

Where: $P_b = \text{brake power, kW}$

$T_b = \text{engine brake torque, N} \cdot \text{m}$

$N_e = \text{engine speed, RPM}$

3.4.1.2 Calculation of Specific Fuel Consumption

When assessing the performance of a piece of agricultural equipment, or any other engine based machine, it is important to include the ability of that machine to convert the fuel into useable work. A commonly used metric for the ability of an engine to convert fuel to work is the specific fuel consumption (SFC) of the engine. This metric is not constant with all speeds, torques, or fuel flows for an engine. Instead, this metric includes the combined effect of speed, torque and fuel flow. The most efficient operating point for a machine on the basis of fuel use to power output is the point at which SFC reaches a minimum according to the relationship (Srivastava et al. 2006)

$$SFC = \frac{\dot{m}}{p} \tag{2}$$

Where: *SFC* = specific fuel consumption, fuel consumption rate per power

$\dot{m} =$ fuel consumption rate

$P =$ power

3.4.1.3 Draft and Power Requirements

The draft force and power of a tool, typically tillage or seeding equipment, is defined as the force required to pull that tool at a given speed and the power that is required to maintain the given travel speed. The draft and power are important parameters in determining the size of tractor required to pull the tool to ensure that a tractor is not undersized, as it would not be able to pull the tool, or oversized, as the tractor would be capable of producing a much greater amount of power than required of the tool. The relationship determining draft power is represented by (ASABE D497.7 2011)

$$D = F_i[A + B(S) + C(S)^2]WT \quad (3)$$

Where: D = implement draft, N (lbf)

F = dimensionless soil texture adjustment parameter from D497.7 Table 1

i = 1 for fine, 2 for medium, and 3 for coarse textured soils

A, B, C = machine specific parameters from D497.7 Table 1

S = field speed, km/h (mile/h)

W = machine width, m (ft) or number of rows/tools from D497.7 Table 1

T = tillage depth, cm (in) for major tools, 1 (dimensionless) for minor tillage tools and seeding implements

3.4.2 Field Performance Research

The relationships shown in theory based equations are important to remember when analyzing machine performance, but the analysis of machine performance in more detail requires the researcher to search for evidence that more than verifies or disproves these relationships. The analysis of machine performance should include the visualization of machine variables in order to observe relationships between multiple sources of data that may not strictly follow theory. In the study of in-field performance of an agricultural tractor by Yule et al. (1999), a data acquisition system was installed with various sensors to record variables regarding the tractor's performance. The following variables were collected from the tractor:

- Engine Speed
- Torque
- Fuel Consumption
- Power

- Specific Fuel Consumption
- Engine Utilization
- Theoretical Speed
- Actual Speed
- Wheel Slip
- Field Slope

From these collected data, the slope of the field was determined to have a very significant influence on wheel slip. It was observed that the northwestern and southeastern portions of a particular field of interest corresponded to the areas of both high wheel slip and high slope. The information regarding the presence of higher wheel slip in areas of high slope was not unexpected. However, the results of the study indicated that the setup of the machine was not optimal for the conditions of the field (Yule et al. 1999). To make conclusions about the performance of agricultural machinery, visual mapping of machine parameters was a useful tool, but human inspection cannot always draw meaningful conclusions from maps alone. Through the use of descriptive statistics including the mean, median, mode and standard deviation, areas of operation that are outside the practical range of a particular machine and operator operation were identified (Yahya et al. 2009; Singh and Singh 2011). Through the use of descriptive statistics and visual mapping of machine parameters, it was found that farmers used, on average, only 60% of their tractor's rated engine capacity. Additionally, it was shown that through the optimal use of the Shift-Up, Throttle-Back (SUTB) principle, farmers could save between 11.3 to 20% of fuel from knowing where their operation was least efficient (Grogan et al. 1987).

In order to provide the machine operator and the farmer with additional information, Boon et al. (2005) used the theoretical relationships between machine parameters in addition to the record of historical machine parameters to provide both an analysis of the machine's performance in the field and a prediction of the machine's future performance along with the farmer's future need. The use of an automated soil penetrometer-shearometer was used to

determine the penetration resistance and shear stress in a field. Using established relationships between penetrative resistance and soil shear stress, the draft force required for tillage was calculated. From the draft force, travel speed and the fuel consumption were subsequently calculated for the power availability of the tractor and mapped on a geographic basis within the field. These values provided the research team with a theoretical number to which the machine's actual performance could be compared. Based upon the maximum draft force and required drawbar power at common operational speeds, predictive measures for future tractor and implement requirements were made (Boon et al. 2005).

3.5 Summary of Literature Review

The literature reviewed in this study covered academic and professional fields relevant to the four major topics of the study: dealing with complex data, evaluating data quality, processing data in ArcGIS™, and evaluating the performance of machines. The general consensus among the reviewed authors whose works discussed complex data and data quality was very general. All authors indicated what they found to work best for their particular application, but they cautioned future researchers from blindly applying their strategies for processing data and assessing data quality. The common recommendation provided by these authors was to include expert analysis of the data analysis process and the evaluation of data quality. This expert analysis should be applied at the initiation of a project through the creation of a plan that includes determination of data quality, acceptable levels of quality, accountability for different aspects of quality, and any other project specific items.

The literature reviewed regarding ArcGIS™ and machine performance evaluation was much more specific in guiding future research. The information describing ArcGIS™ operations provided suggestions regarding the capabilities of the program. Additionally, these studies

provided suggestions for the geospatial mapping, statistical analysis, and analysis of usability of geospatial information taken from a third party. The sources describing machine performance provided two distinct approaches. The theoretical sources provided a solid foundation on which the performance of a machine could be compared to the ideal operation. The practical sources and studies with applied theory and practice provided the suggestion that theory based metrics can be used as a comparative measure for practical operations of machines, but they also suggested that machine performance data should not be expected to strictly follow trends expected by theory. Therefore, the practical analysis of machine performance data should remain unbiased and attempt to discover unexpected trends in the data.

CHAPTER 4: MATERIALS AND METHODS

The ability to research the performance of agricultural machines during annual operations was a difficult task due to the variability of environmental factors and influence of different hybrid types on yield. The methods used in this study were intended to retain the strengths of studies that have been published with the same objectives while avoiding their flaws. Section 4.1 describes the data collection devices that were used to collect performance data in this study. Section 4.2 describes the farms from which the data were collected along with the different types of machines that generated the data. The preprocessing tasks that were required for the data before it could be analyzed with ArcGIS™ (ArcGIS™ for Desktop 10.1, Esri 2012) are included within Section 4.3 along with an analysis of the data quality. Finally, a description of the data importation process and a description of how the data were analyzed in ArcGIS™ to determine the quality of machine performance through the farms are included in Section 4.4. A flow chart describing the general process of data collection and processing is included in Figures 1a and 1b for the data associated with the MTG Data Logger and the StarFire™ and GreenStar™ sources respectively.

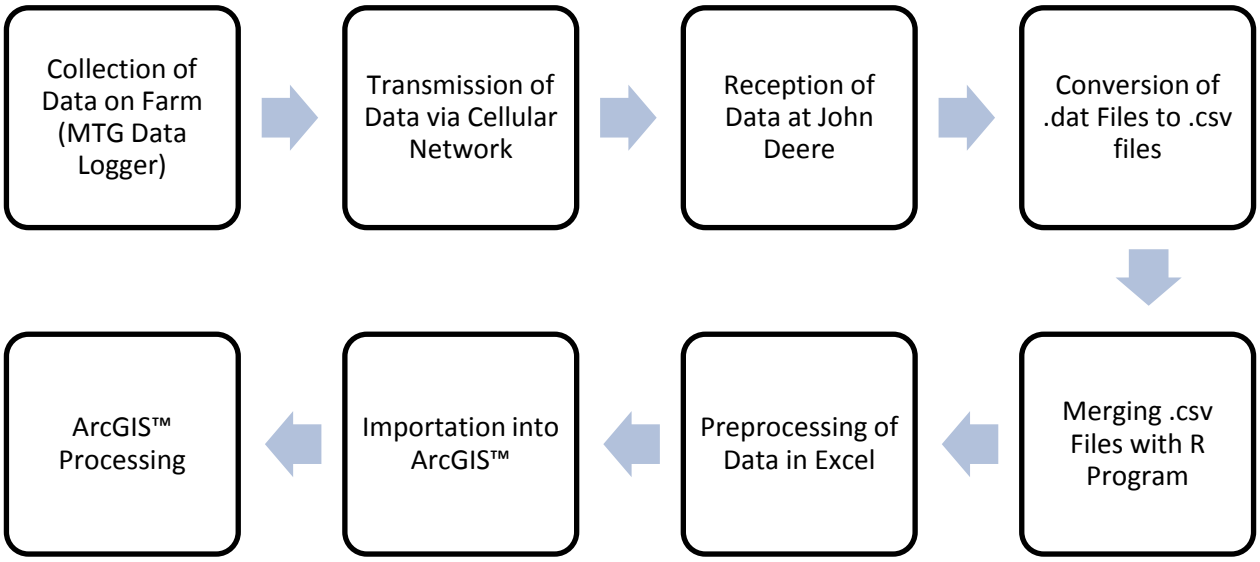


Figure 1a – Data process flow chart for MTG Data Logger

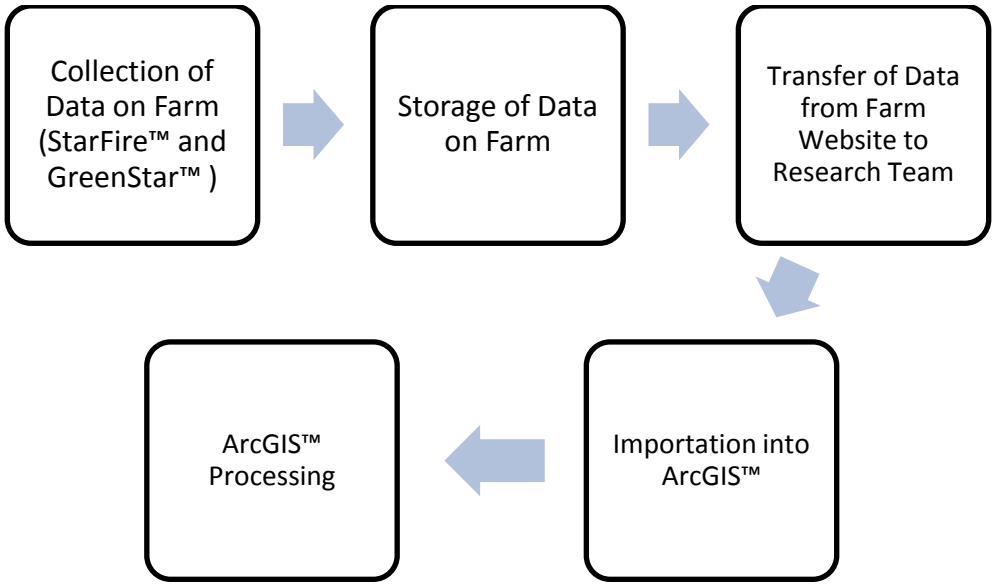


Figure 1b – Data process flow chart for StarFire™ and GreenStar™

4.1 Data Collection Devices

The data that were collected from the machines were recorded by proprietary devices manufactured by John Deere for consumer use that have been adapted to the specific requirements of this project. These devices allowed data to be collected from operations throughout the season without the interference of scientists in the daily operations on the farms. This aspect of the data collection attempted to create the least biased set of data possible while still ensuring the completeness of the data. In addition to the description of the data collection devices used in this study, a description of the amount of data collected via each method is provided in section 4.1.3.

4.1.1 John Deere GreenStar™ Display and StarFire™ Receiver

The John Deere GreenStar™ Display and the StarFire™ Receiver are an integral part of John Deere's precision agriculture system. The StarFire™ Receiver refers specifically to the GPS receiver that was used to determine the position of the machine. The StarFire™ Receiver then transmitted information about the machine's position to the GreenStar™ Display. The GreenStar™ Display was responsible for collecting other relevant information from the machine and recording that information along with position. This information was later exported in the form of Esri shapefiles to be viewed and or analyzed in ArcGIS™ (John Deere 2013a; John Deere 2013b).

The John Deere StarFire™ Receiver is the line of GPS receivers that are available to use in conjunction with the GreenStar™ Display. These receivers are available in three technologies that provide varying levels of accuracy: SF1, SF2, and RTK. These technologies provide the GreenStar™ Display with position information that is accurate to within plus or minus 23 cm (9

inches), 5 cm (2 inches), and 2.5 cm (1 inch) respectively. The StarFire™ Receivers were specifically designed to provide the user and the GreenStar™ Display with terrain compensation. The terrain compensation includes three subcomponents for rotations of the machine about its three axes resulting in changes in roll, pitch, and yaw. In order to correct for these factors, the operator enters the specific location of the StarFire™ Receiver into the GreenStar™ Display along with other parameters regarding the machine's setup (John Deere 2013a).

The GreenStar™ Display is the computer for the precision agriculture functions for the John Deere machines. It can record parameters on the machine for later interpretation including: jobs, field name, variable-rate applications, coverage maps, conditions, geographic location, yield, and moisture content. The GreenStar™ Display can also export the information collected in the form of shapefiles. These shapefiles can be saved to a USB storage device in order to transfer them to a computer for use in Esri's ArcGIS™ software. The GreenStar™ Display and the StarFire™ Receiver work in tandem to record most information that is relevant to farmers and provide that information in an easy to use and interpret format. This setup, however, does not include all parameters that were relevant to the study of machine performance. The GreenStar™ display has a model dependent amount of on board memory with the capability of being expanded with USB flash memory devices (John Deere 2013a; John Deere 2013b).

4.1.2 John Deere MTG Data Logger

In order to fully analyze the performance of agricultural machines in this study, certain messages were recorded from the machine's CANBUS. To record these messages, Modular Telematics Gateway (MTG) Data Loggers were installed on the machines. This data logger originated within the construction division of John Deere as a major component of the JDLink™ system. JDLink™ was designed to provide operation owners and managers with the location,

operating characteristics, and other pieces of information regarding their fleet of equipment. The MTG Data Logger collects its own position information through an internal GPS receiver. This position is not as accurate or precise as the GPS position determined by the StarFire™ receiver due to the desired objectives with collecting position for operation owners and managers' use. The StarFire™ GPS location must be both accurate and precise in order to provide a consistent reference for precision agriculture operations, and the position from the MTG Data Logger is used primarily to inform the general position of the machine(s) to the operation owner or manager. This information is transmitted wirelessly from the machine to the operation owner or manager to a variety of devices including smart phones, desktop computers, and laptop computers (John Deere 2013c). JDLink™ was adapted to the agriculture industry to provide similar information regarding agricultural equipment (John Deere 2013d). Specifically of interest to this study was the MTG Data Logger's ability to collect fuel consumption, engine speed, engine torque, transmission settings, various indicators of machine speed, and other operational parameters. This information was collected and stored on the MTG Data Logger and transmitted via cellular tower to a central location where the accumulation of data from all machines relevant to that operation occurred (John Deere 2013c; John Deere 2013d).

4.1.3 Data Collected by Device

The data used in this study were collected via two distinct methods. The data from the StarFire™ and the GreenStar™ system were collected during the normal occurrence of operations on the farms. These farms used this system to collect information regarding farming operations for their own use. These data were relayed directly to the research teams from the farms with little involvement from John Deere. The data from the MTG Data Logger were collected and transmitted to the coordinating research teams by John Deere. The only portion of

the data that came directly from the farms was limited to only the wheat harvest on Farm 2. The vast majority of the data was collected by the MTG Data Logger.

4.2 Farms and Equipment

Section 4.2 contains a description of the farms from which the data were taken during this study. It additionally contains model descriptions of the equipment that each farm used in their operations. Specific information regarding the identity of each farm is not provided due to client confidentiality.

4.2.1 Farm Details

In order to evaluate the performance of machines in agricultural operations, John Deere contracted Midwestern farms to participate in this study. The first farm, subsequently referred to as Farm 1, was an approximately 5,600 hectare (14,000 acre) farm located in the east-central portion of Iowa. Farm 1 was considered a large-scale corn growing operation that used conventional farming practices. The second farm, hereinafter referred to as Farm 2, was located in the western portion of Kansas. Unlike Farm 1, Farm 2 grew a diverse range of crops including: winter wheat, corn, grain sorghum and some alfalfa. With approximately 4,500 hectares (11,000 acres), Farm 2 was of a similar size to Farm 1. Farm 2 differed from Farm 1 in both the type of farming practices and the type of land. Instead of conventional farming practices as performed on Farm 1, Farm 2 utilized a no-till strategy. This no-till strategy differs from conventional farming practices in that it does not use tillage to work the land in between crops. This required different spraying practices of herbicides, pesticides, and fertilizer than conventional farming practices. The land of Farm 2 also differed vastly from Farm 1. Farm 2 contained many fields with terraces for water retention and erosion control purposes.

Additionally, some fields within Farm 2 were irregularly shaped due to the presence of irrigation systems. Most fields within Farm 1 had neither terraces nor irrigation systems during this season.

4.2.2 Equipment

Both Farm 1 and Farm 2 operated entirely on John Deere equipment during the duration of this study. The only exception to the use of John Deere equipment was the use of pickup trucks, semi-trucks, grain carts, and other equipment that John Deere did not manufacture. For the 2012 season, Farm 1 operated one John Deere 4940 sprayer, two John Deere 6170R tractors, five John Deere 8360R tractors, five John Deere 9460RT tractors, three John Deere S680 combines, and one John Deere 9870 combine. For the 2012 season, Farm 2 operated one John Deere 4830 sprayer, two John Deere 8345R tractors, and three 9770 combines.

4.3 Data Preprocessing

The raw data collected from the MTG Data Loggers on the machines that were studied were not in a format in which they could be directly processed with ArcGIS™. As a result, the data had to be preprocessed in order to change the format of the data into one that could be imported into ArcGIS™. The data was originally created in data files (.dat) that were converted to comma separated value files (.csv) by a utility created by researchers from Purdue University who were contributing partners to this study. The copyright for this program has been retained by its creators and is, therefore, not included in this publication. Once the files were converted to .csv files, they were merged to create one summary file per machine. Due to file formatting issues, the files were manually sorted for error values within Microsoft Excel (Version 14.0.6129.5000, 64-bit, Part of Microsoft Office Professional Plus © 2010 Microsoft Corporation). Finally, the raw data were manually inspected to determine the quality of the data

before proceeding with the analysis. The measure of quality in this project was determined based primarily on the validity of the GPS locations in reference to the known GPS positions of the farms from which the data were collected.

4.3.1 Data Processing with R Studio

Each file originating from the MTG Data Loggers represented one instance of the machine being turned on. Therefore, the number of files for a given machine in a season was representative of the number of times that particular machine was started during that season. This characteristic of the data resulted in the creation of a large number of files for each machine. In order to organize the data better and provide metrics for the farm as a whole, these files were merged for each machine. However, it was not possible to create only one file for each machine. The number of parameters that were recorded by the data loggers was changed multiple times, in some cases, for each machine during the season. To compensate for the inability to merge .csv files of different column order and lengths, the files for any particular machine were merged according to file setup. The resulting files, between one and three per machine, allowed the machines' performance to be compared both on a field-by-field basis and to the average operation of that machine across all fields. In order to merge the files together, code was written by Andy Stevens, from John Deere, in R programming language and run in RStudio (Version 0.97.236 © 2009-2012 RStudio, Inc.). The text of the code is reproduced in Appendix A with permission from Andy Stevens.

4.3.2 Data Processing with Microsoft Excel

In an attempt to clean the data of corrupt values, the data was manually sorted and inspected within Microsoft Excel. Using the sort function within Excel, the values for latitude

and longitude were sorted from largest to smallest. This grouped all null values for the latitude or longitude at the top or bottom of the spreadsheet with the plausible values centered in the middle. These null values were removed from the data along with any values of latitude or longitude that were significantly out of the range of values expected for the location of each of the farms. The main objective of this process was to eliminate the data points that did not have any valid geographic reference.

In addition to removing non-georeferenced points, machine parameters that had the same values throughout the machine's entire operation, referred to as flatlined parameters due to their constant nature, or those parameters for which no value was recorded through the entire machine's operation were removed from the data set. The removal of non-georeferenced data points and non-meaningful records of machine parameters was performed in order to reduce the time required to process the data, reduce the computer processing ability that was required for each operation, and to only include data points that accurately represented the machine's operation in the final results. This operation would have been performed via an automated program had the same parameters been recorded in the same order across all machines of the same type. In order to automatically remove non-georeferenced data points or constant value parameters with the current file setup, the program would have had to be rewritten for each file setup.

4.3.3 Data Quality Analysis

The analysis of the quality of the data for this study was a difficult task due to the lack of a known standard for each machine, operator, and farm to which the collected data could be compared. Due to this difficulty, the five components of describing geospatial data, logical consistency, positional accuracy, temporal accuracy, attribute accuracy, and completeness of data

(Devillers et al. 2007), could not all be analyzed. Therefore, a new metric to determine data quality relevant to this study was created. This metric, "Percent Data Lost," is best described as the percentage of data points collected that were not within the expected range of values for each parameter. To determine this percentage, the size of the .csv file was compared before and after the removal of null values and flatlined parameters as described in Section 4.3.2. "Data Loss" does not imply that all of the remaining data was definitively accurate. Instead, this metric was intended to indicate the percentage of the data that was collected that was not contained within the expected ranges of each farm's geographic coordinates. This indicates that the actual percentage of data that was an accurate representation of the operations can be no greater than the percentage of data that was within the expected ranges. In fact, the percentage of data that was an accurate representation of the machines' actual performance was expected to be much less than the percentage of data that was in the plausible range due to possible mismembership of the data.

As a supplement to the numeric analysis of data quality, the data were inspected on a qualitative basis for completeness. Each field was viewed with the data points for each machine plotted across the field. This method was used to determine that there were no gaps in the record of a machine's operation in the field. This evaluation method led to discoveries that some machines and fields showed inconsistent data collection or lack of data collection entirely in some portions of the field. The purpose of visually inspecting the data within ArcGIS™ was to satisfy the suggestion of several publishers of data quality research that indicated the use of expert judgment as one of the most reliable forms of determining and quantifying the quality of any given data set.

4.4 Data Processing in ArcGIS™

The following section contains information relating to the use of ArcGIS™ in the processing of information in this study. This section is divided into the data importation and data analysis with detail regarding each step related to the corresponding topic.

4.4.1 Data Importation

The first step in analyzing information in ArcGIS™ was the proper importation of the data within ArcGIS™. In order to ensure that the data being imported to ArcGIS™ appeared in the correct location, basemaps were added so that alignment could be verified visually. These basemaps were added to the document by using the dropdown menu shown in Figure 2. The two maps that were used consistently in this study were the "Imagery" and "Streets" basemaps from ArcGIS™ online. The "Streets" basemap provided a visual reference to determine if the fields were in the proper locations in reference to the states and cities in which the farms are located. The "Imagery" basemap was included to reference the shape of the fields from aerial imagery to the shape of the fields' polygons and the generic shape of the machines' paths within the fields.

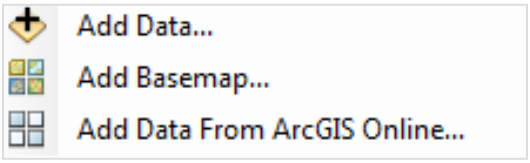


Figure 2 – "Add Data" menu within ArcGIS™

To add the data collected from the machines in the study to the document, the "Add Data…" option from the "Add Data" dropdown menu shown in Figure 2 was selected. The .csv files were then selected by location on the computer's hard drive and imported into the document. At this stage of the importation process, the data did not appear geographically

referenced. Instead, they were available in ArcGIS™ in the spreadsheet format only. To align these files geographically, the user right-clicked on the file and chose the option "Display XY Data…" from the menu that appeared (As shown in Figure 3).

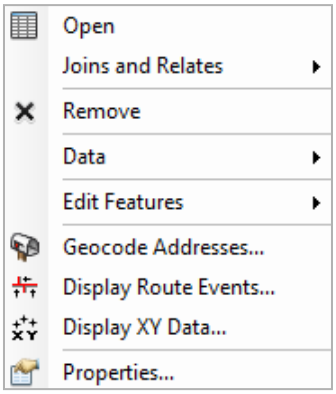


Figure 3 – Select "Display XY Data…" to choose position parameters for data set.

This option displayed a menu in which the user specified the field that indicates the X, Y, and optional Z values that indicate the point(s) location. In the case of this study, the X and Y variables were longitude and latitude respectively. This information originated from either the StarFire™ or the MTG Data Logger's GPS receivers. The name for the field(s) with the latitude and longitude varied based on the specific file's configuration. In the file being georeferenced in Figure 4, the X and Y fields are named "Lon" and "Lat" respectively. In order to provide the proper scale with which the coordinates will be projected, the user must also specify the "Coordinate System of Input Coordinates," also shown in Figure 4. The coordinate system must be specified because the world is an irregular shape. Various coordinate systems exist in order to project coordinates collected from the roughly spherical shape of the earth to a two-dimensional representation.

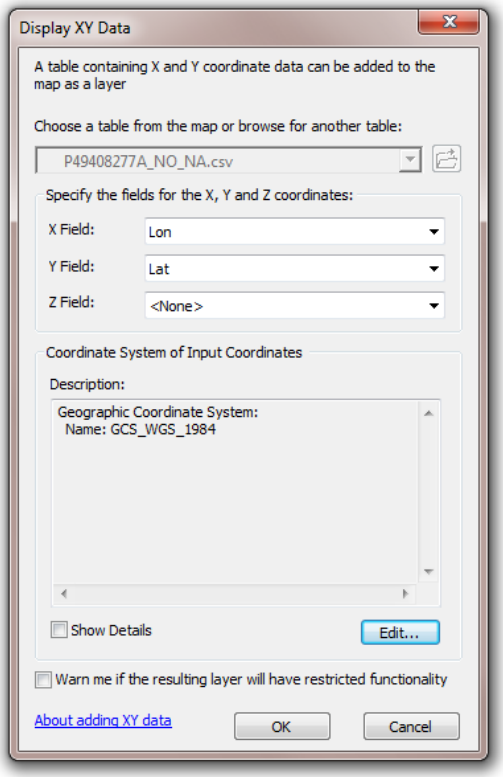


Figure 4 – Information required to georeference the data in the "Display XY Data" menu

Although this method allowed the user to import the data and view it within ArcGIS™, this did not actually save the .csv as a layer file, the preferred file format for ArcGIS™. To enable faster data processing and editing features for ArcGIS™, the data was exported as a layer file, and then reimported into ArcGIS™ in the new file type. This was accomplished by right-clicking on the specific file name, selecting the "Data" option, and then the "Export Data…" option in the submenu as shown in Figure 5. After the data set was reimported into ArcGIS™ in the new format, it was ready to be analyzed according to the objectives of the project.

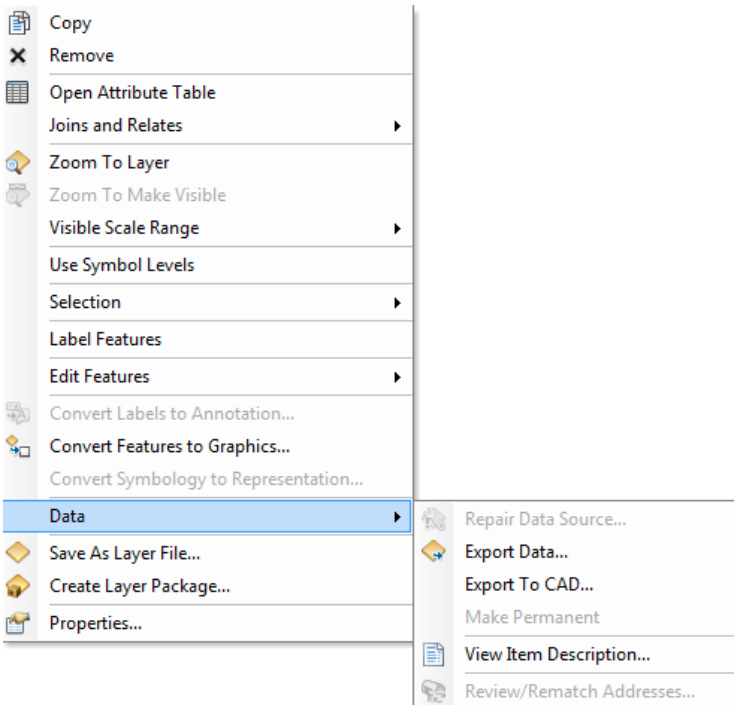


Figure 5 – Process for exporting .csv files as layer files in ArcGIS™

4.4.2 Data Analysis

Two approaches to the analysis of the data collected in this study were taken. A qualitative approach was taken through the visual inspection of the data by plotting the parameters' change throughout the machines' operations in the field. The second approach to analyzing the data was a quantitative approach. This was intended to both confirm the trends in data that were discovered through the visual inspection of the data and to find relationships between machine parameters that may have been missed through visual inspection alone.

4.4.2.1 Qualitative approach

In order to avoid focusing only on expected results in the analysis of the machine performance data, the data were visually inspected to look for spatial and temporal trends that were not anticipated. In order to visualize the information in a way that allowed trends to be

seen, the ability of ArcGIS™ to map variables with differing symbol characteristics was utilized. In the "Layer Property" menu, under the "Symbology" tab, the characteristics of the symbol representing each point of the data were changed based on the value of a parameter at that point. These characteristics include the shape, size, and color of the symbol. Examples for the process of defining the symbol characteristics and the resulting output for those specifications are included in Figures 6 and 7 respectively.

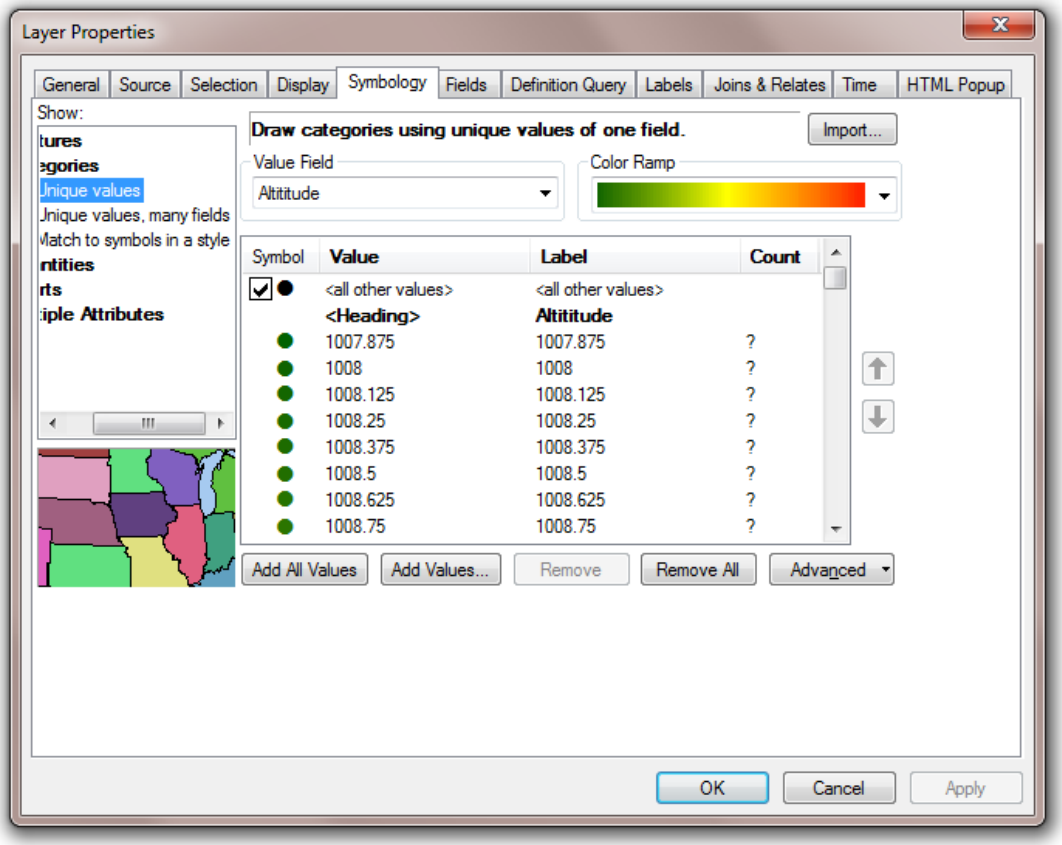


Figure 6 – Defining symbol characteristics for varying values for the field "Altitude"

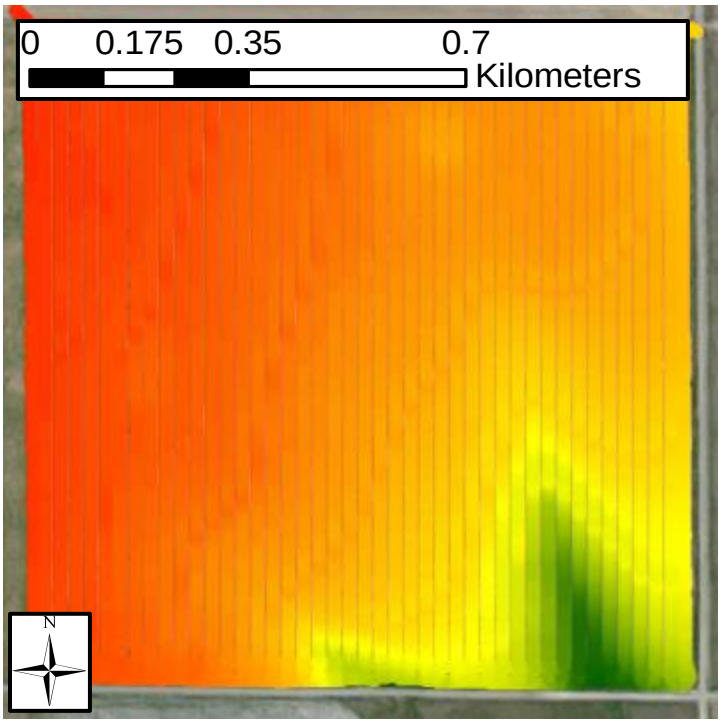


Figure 7 – The plot of "Altitude" with the use of variable symbol characteristics dependent on parameter value with a continuous color variation from green in areas of low elevation to red in areas of high elevation

The application of varied symbol characteristics was not only useful in the continuous mapping of machine parameters. It was also useful in applying operational rules to the machine. Figure 8 shows the application of machine operational state rules to the use of varied symbol characteristics. The red data points indicate points that have no motion or motion that was slower than those expected of the machine during operation. The blue data points were scaled to include machine operational speeds and are color variable within the operational speed range. The yellow to pink points indicate the data points where the machine was being operated at transportation speeds with a continuous color transition from speeds in the low transportation range to high transportation range having colors yellow and pink respectively. This type of plot was useful in determining locations where the machine was in its idle state for the greatest

amount of time or locations where the machine was being operated either faster or slower than the farm managers' target speeds within the field.

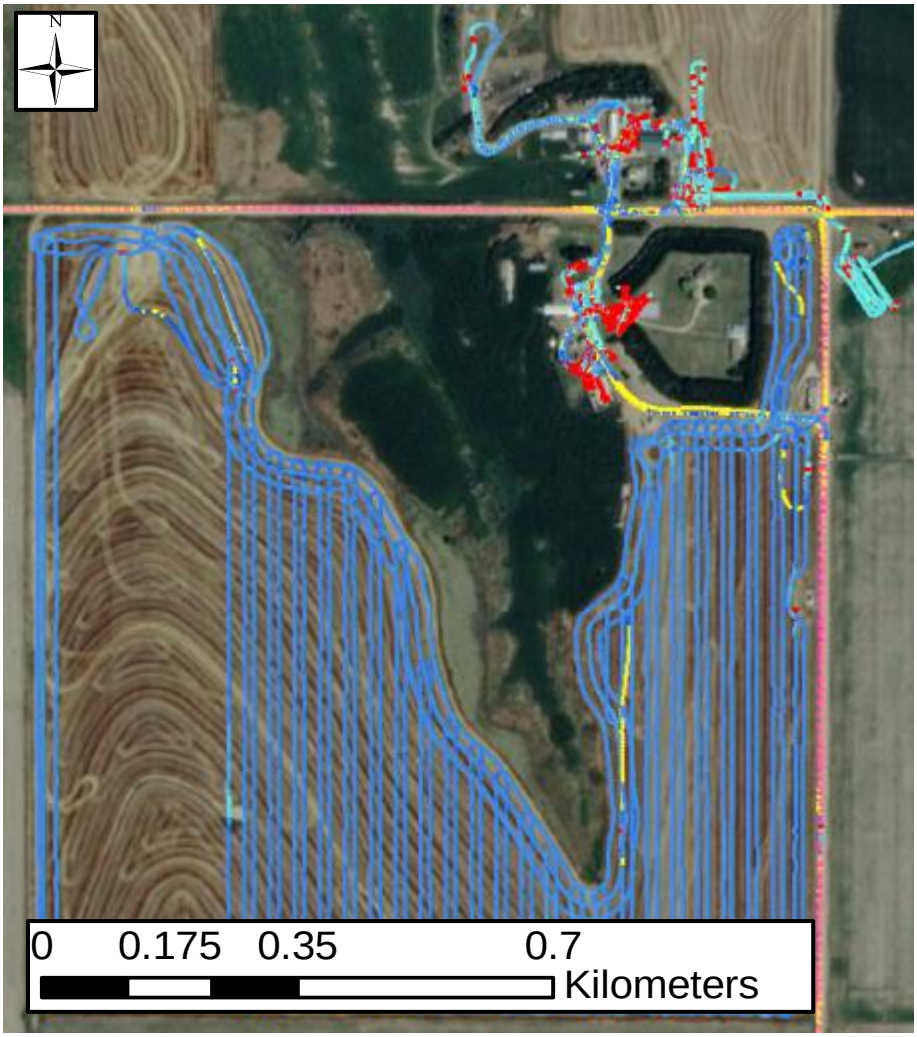


Figure 8 – Application of machine operation rules to varied symbol characteristics with red points, blue points, and yellow to pink points indicating idle, operational and transportation speeds, respectively

4.4.2.2 Quantitative Approach

Although the observation of the change in parameter values throughout the machine's operation provided useful information from which researchers or farmers could draw conclusions, this type of data analysis provided no quantitative terms in which correlations could

be made. For example, if a geographic map of fuel consumption is created to compare to the map of elevation as depicted in Figure 7, the fuel consumption map may show increased fuel consumption values as the machine was traveling up the hill and decreased fuel consumption as the machine was traveling down the hill, but this does not prove a statistically significant correlation between the two variables. Instead, the use of qualitative investigation of the relationships between variables provides an indication as to which relationships should be explored more thoroughly through statistical means.

In order to determine the statistical relationship between any two or more variables, the ArcGIS™ tool "Ordinary Least Squares" found in the ArcToolbox under "Modeling Spatial Relationships" in "Spatial Statistics Tools" was used. The "Ordinary Least Squares" tool provides a simple linear regression or a multivariate linear regression analysis of the data based on the number of expected explanatory variables for the dependent variable. To perform this analysis on a given set of data, the layer file that contained the data to be inspected was specified in the "Input Feature Class" section of the menu shown in Figure 9. The "Unique ID Field" was a column within the spreadsheet form of the data that contained a unique value for every data point collected within the set. This gave the computer instruction as to which dependent variable corresponded to a respective input independent variable. The "Output Feature Class" was the specified location of the output layer file that would contain color coded information regarding the standard deviation of any particular data point. The "Dependent Variable" was the variable whose association to another variable was being determined. The "Explanatory Variables" section allowed the user to specify one or more variables that were expected to explain the trends of the "Dependent Variable" being examined by the tool.

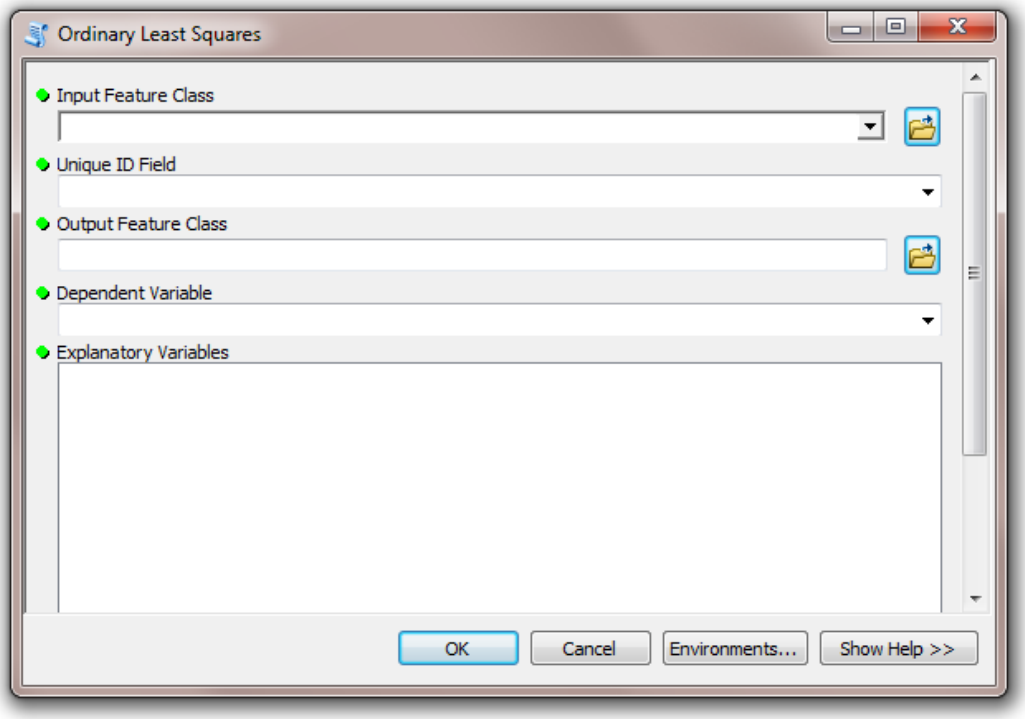


Figure 9 – The "Ordinary Least Squares" menu for statistical analysis of data

The results of data analysis with this tool were provided by ArcGIS™ in the form of a .pdf file. These .pdf files contained information regarding the linear model that was generated by the tool including the intercept and slope of the dependent variables that were analyzed. The output also included a t-Statistic and corresponding probability value that indicated the statistical significance of the model stating that the two parameters being analyzed are in fact related in a statistically significant way. Histograms and scatterplots for each explanatory variable and the dependent variable were plotted to show how each variable was distributed and how the plot of each variable appeared. For the verification or further inspection of the model, a plot of the residuals vs. the plot of predicted values of the model, a histogram of the residuals, and other statistical metrics were provided.

These methods of analyzing the performance of agricultural equipment were useful in determining the operating characteristics of machines and the interdependence between those characteristics. They did not, however, present farm operators with a comparison amongst fields. As such, the impact of field characteristics from slope, size, and shape were also considered in the inspection of machine operation. To accomplish this task, generic statistics regarding fuel consumption, average speed, and other variables were generated on both a field-by-field basis and a farm-wide basis. The statistics for the mean and standard deviation of these variables indicated the fields that had parameters with higher standard deviations, showing less consistent operation, and extreme values for parameter means, showing fields that were above or below farm averages. These metrics were intended to assist farm operators in increasing the efficiency of their operation by selecting better field characteristics when determining fields to rent or purchase for future seasons.

CHAPTER 5: RESULTS AND DISCUSSION

The sections contained within this chapter of the study present the results of the analysis of the data as described in the previous chapter. The first, and most important, analysis of the data is presented in Section 5.1 with the assessment of the quality of the data. The results of the visual analysis of the data are presented in Section 5.2 along with hypotheses that should be further inspected through the use of statistical methods for confirmation or contradiction. Section 5.3 contains the statistical metrics describing the statistical correlation between variables discovered in Section 5.2 and variables that were not observed to show significant relationships through visual inspection alone.

5.1 Data Quality

The quality of the data was analyzed through two independent methods. The first of the two methods was a quantitative measure of the data that were collected and had no accurate geographic position associated with them. The second method was a qualitative analysis using expert judgment to determine what portion of the total was actually collected in a complete and accurate fashion. This second method was performed by visually mapping the data and inspecting it for consistency with expected geographic positions of data points. The qualitative analysis of the data quality occurred after the quantitative measure of data quality was performed and after the data cleaning process was completed. After the quantitative and qualitative analyses of data quality were performed, an analysis of the quality of data documentation was carried out.

5.1.1 Quantitative Measure of Data Quality

The quantitative measure of data quality raised serious questions about the validity of any conclusions that could be drawn upon the data and questions about the confidence of those

conclusions. The quantitative measure describing the quality of the data was defined as the percentage of data lost through the data cleansing process as described in Section 4.3.2. This process removed all points that have no geographic reference and the parameters for which no unique values were recorded through the entire operation of a machine for a season. To provide a consistent comparison between the amounts of data lost through lack of accurate position coordinates and the data lost in flatlined parameters, the files were compared on the basis of file size. The results generalized by machine of this analysis are included in Tables 1 and 2 for Farm 1 and Farm 2 respectively. The results by each setup of machine files as described in Section 4.3.1 are included in Appendix B.

Table 1 – Data loss associated with Farm 1 on a per machine basis and a farm-wide basis

Farm 1			
Machine	File Size (kB)		Data Loss
	Raw	Cleaned	(%)
JD 4940	71,812	35,708	50%
JD 6170-1	199,991	93,089	53%
JD 6170-2	147,168	68,890	53%
JD 8360-1	247,492	172,114	30%
JD 8360-2	440,955	272,948	38%
JD 8360-3	261,563	161,797	38%
JD 8360-4	482,481	260,873	46%
JD 8360-5	397,917	147,452	63%
JD 9460-1	425,710	260,447	39%
JD 9460-2	487,235	272,146	44%
JD 9460-3	232,890	97,962	58%
JD 9460-4	177,612	94,116	47%
JD 9460-5	192,174	126,710	34%
JD 9870	51,572	36,708	29%
JD S680-1	238,469	169,503	29%
JD S680-2	267,632	142,741	47%
All Machines	***4,322,673***	***2,413,204***	***Mean (Std. Dev.) 44% (10%)***

Table 2 – Data loss associated with Farm 2 on a per machine basis and a farm-wide basis

Farm 2			
Machine	File Size (kB)		Data Loss
	Raw	Cleaned	(%)
JD 4830	375,758	239,123	36%
JD 8345-1	573,719	381,247	34%
JD 8345-2	301,538	178,257	41%
JD 8345-3	338,954	249,296	26%
JD 9770-1	69,506	47,214	32%
JD 9770-2	195,650	73,602	62%
JD 9770-3	315,346	248,434	21%
All Machines	***2,170,471***	***1,417,173***	***Mean (Std. Dev.) 35% (13%)***

The values for "Data Loss" ranged from around one-fifth of the data to approximately two thirds of the data relating to one specific machine. With average "Data Loss" on Farm 1 and Farm 2 being 44% and 35% per machine respectively, decisions based upon these data come with a very low confidence level. The lack of a full record of field operations implied that the geospatial data was definitely incomplete, but it did not imply that the accuracy of the data that was actually collected had been diminished. This indicated that the portion of data points that were actually recorded would provide a sufficient basis on which to analyze the methods of analysis presented in this study, but it would not be capable of providing sufficient basis on which to provide recommendations for new operations. Additionally, the results of this study along with the evaluation of the data collection methods provided a reference for future data to be compared, and they provided those involved in data collection with feedback vital to the collection of complete and accurate data in the future.

5.1.2 Visual Inspection of Data Quality

In order to further investigate the quality of the data, the data were investigated visually to determine the source(s) of error. If the lack of data quality was caused by the lack of cellular reception, then it would have been expected that the data points that were correctly recorded would be concentrated to one area of the field, and the data points that were not correctly recorded according to position would be concentrated in a separate location. As shown in Figure 10, this was not the case. The data points in this Figure were taken from the JD 4830 sprayer on Farm 2. This field appeared to have a complete collection of data. However, close inspection of the field revealed that when the points were mapped according to ground speed, the distance between points was not directly related to the speeds recorded at those points. With data being collected with a sampling rate of 1 Hertz, the distance between points should have been related to the speed at which the machine was traveling for those two points. The observation of the distance between two consecutive points did not reinforce this fact. With speeds held relatively constant throughout the spraying operation, as reinforced by Figure 10, the space between two consecutive, equal-speed points varied illogically. This supported the conclusion that the points within this field were either recorded inaccurately or that they had been corrupted at some other point in the data analysis process.

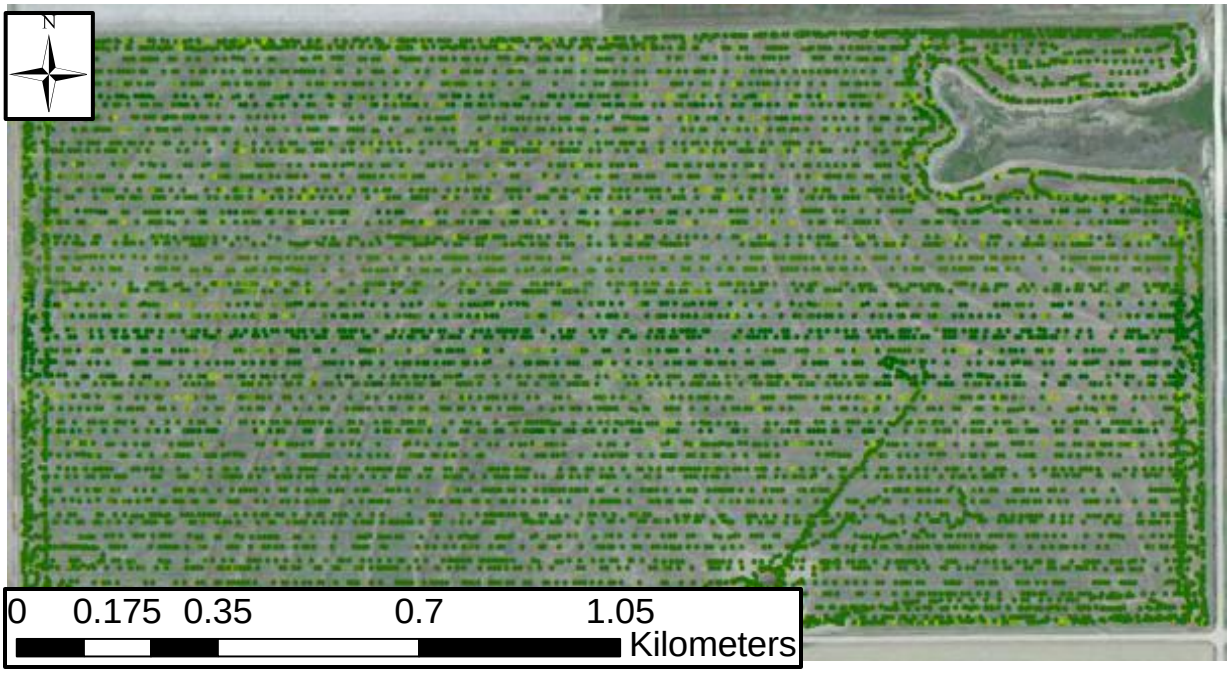


Figure 10 – Plot of a JD 4830 sprayer from Farm 2 in relation to speed with darker green indicating locations of slow speed and lighter green indicating areas of faster speeds

The incomplete record of data points within a field was not constrained to a uniform distribution of unrecorded points within a field. Several instances of groups of missing points were discovered. On Farm 1, the field shown in Figure 11 experienced sections of missing data points along with brief portions of correctly recorded data within the data gaps. These "islands" of properly recorded data within the large gaps in data were easy to identify due to the lack of machine tracks going to or coming from these recorded points. The specific field shown in Figure 11 did not exclusively contain improperly recorded data. There were complete data sets recorded by some machines and incomplete data sets recorded by others, and there were some incomplete data sets with uniformly distributed unrecorded data points similar to what occurred with the JD 4830 sprayer in the field shown in Figure 10.

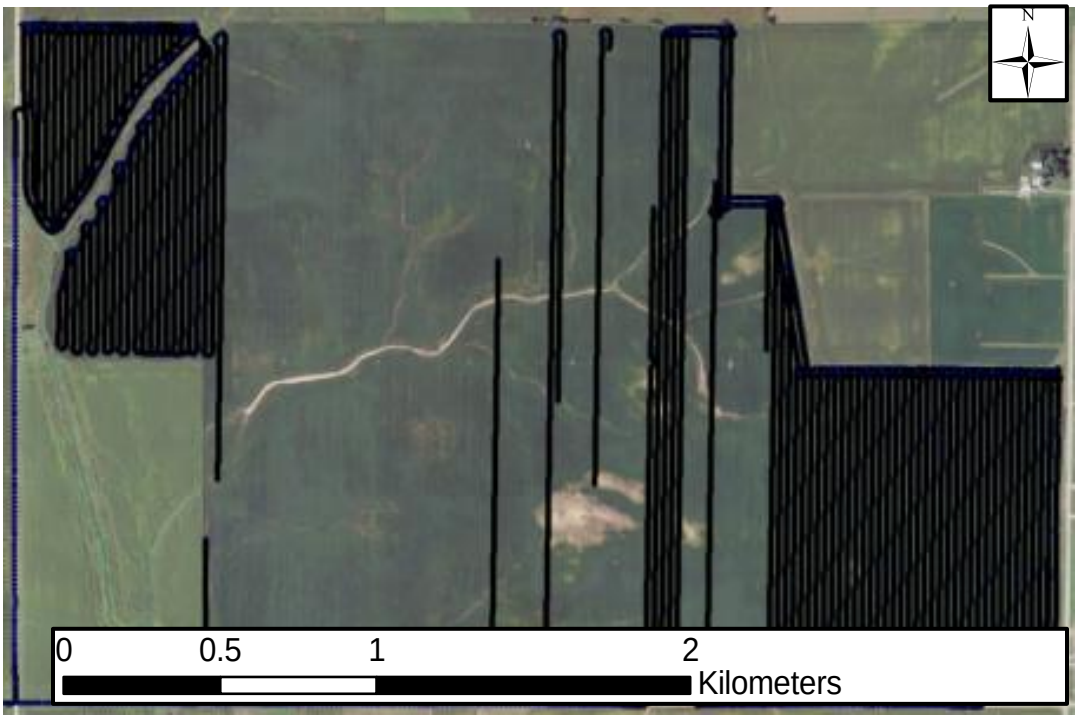


Figure 11 – Large sections of data not recorded with sporadic "islands" of correctly recorded data points in a field from Farm 1

In addition to the completeness of the data, the positional accuracy was determined using a visual analysis. Figure 12 shows data collected during the wheat harvest on Farm 2. The geographic position of the black points was taken from the MTG Data Logger, and the geographic position of the pink points was taken from the StarFire™ receiver and recorded with the GreenStar™ display. The difference in the track of the machine recorded with each of the two different types of GPS receiver indicated the difference in the accuracy of each GPS technology. As apparent in Figure 12, the pink tracks are much smoother and indicate what would be expected of the actual track of the machine. The black tracks, however, deviate from what would be the expected path of the machine with varying magnitude.

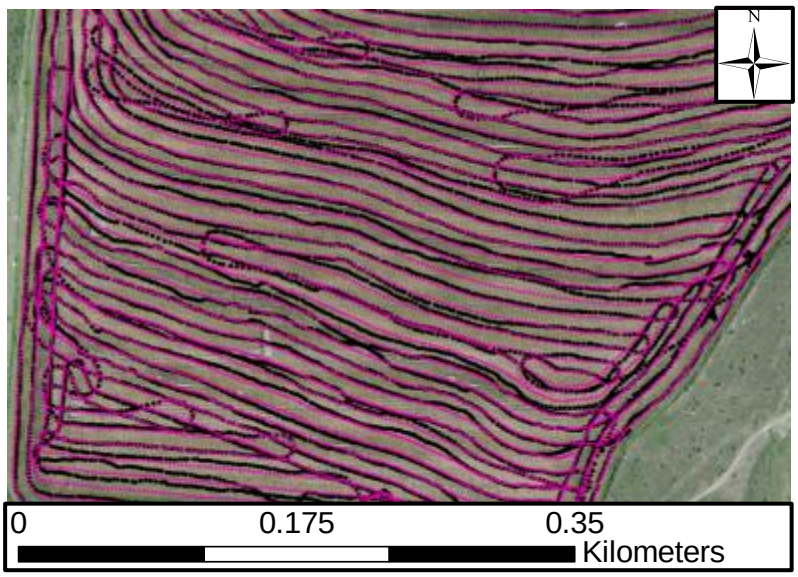


Figure 12 – Plotted difference in accuracy and precision of location recorded by the StarFire™ GPS Receiver (Pink) and the MTG Data Logger GPS Receiver (Black)

5.1.3 Data Documentation

After analyzing the quality of the data quantitatively and qualitatively, portions of the data were determined to be of significantly poor quality as determined by the large number of points with invalid GPS coordinates. However, there were enough fields and machines for which data appeared complete enough to infer meaningful information about trends in machine parameters. The documentation of these data was not complete enough to allow full analysis of these parameters. The files containing the data included the name of the parameters being collected along with the values for those parameters. These files did not contain proper descriptions of what the parameters were. From the name of some parameters and the units of the parameters given in the setup file for the MTG Data Logger, the meaning of some of the parameters could be inferred. For example, "CAN_FuelRate" with the units of "gal/hr" could be reasonably inferred to be the rate at which the engine was consuming fuel in gallons per hour. Some parameters were not as easy to determine their meaning. Attempts to determine accurate descriptions of the parameters that were recorded proved unsuccessful. A data dictionary for

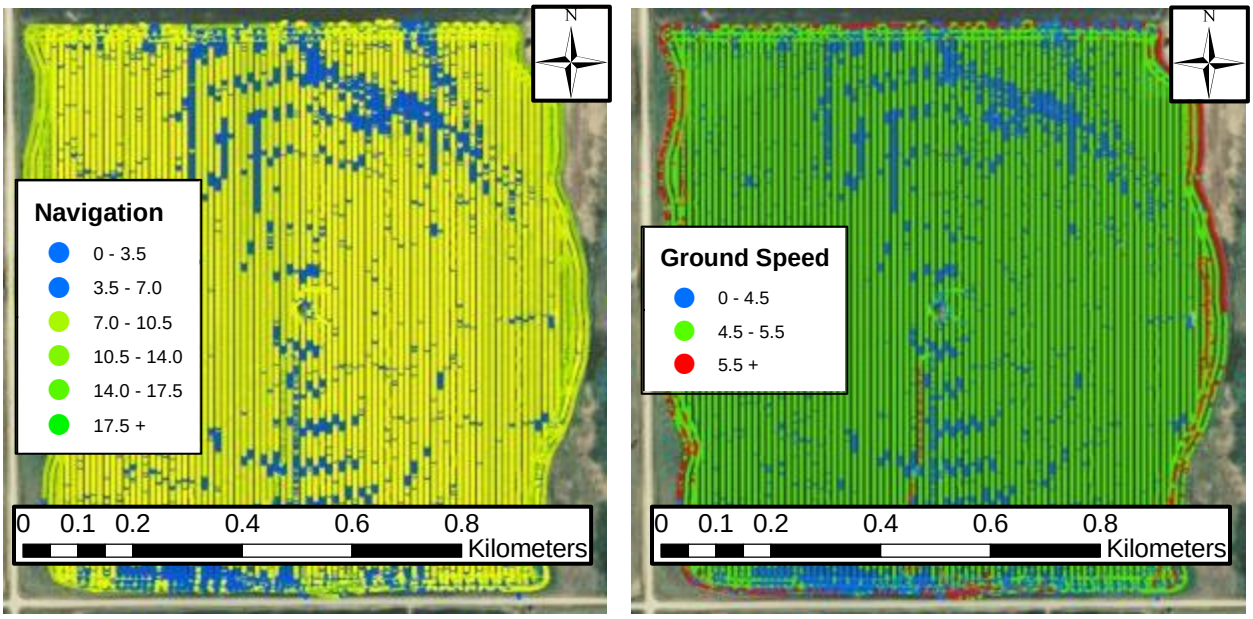
parameter names according to the name within the machine computer had been created when the machines were designed, but documentation of the changes in names, units, and other descriptions had not been updated with every change in model year. Additionally, the name of a parameter might have been the same for a combine and a tractor, but the actual parameter that was being recorded or the units of that parameter could be different. The lack of proper documentation regarding the parameters being collected decreased the value of the data in such a way that the confidence of any conclusions based on this data was very low. The data in this study were best used to assess the process of analyzing machine performance and to provide suggestions as to the improvement of future machine performance data collection. The units for any parameter that are not clearly defined in this document can be assumed to be undocumented.

5.2 Visual Analysis of Data

Visual analysis of the data was performed in order to provide an initial qualitative analysis of the data to assess overall machine operation characteristics, the relationships between different parameters of the same machine, and the relationship between the same parameters of different machines. The analysis of overall machine operation characteristics was performed on a farm-wide basis to determine metrics for the division between states of the machine: idle, in-field operation, and transportation. The analyses of relationships between different parameters of the same machine and of relationships between the same parameter on different machines were performed on a field by field basis in order to determine the influences of one machine parameter on another and the influences of field characteristics on the operating characteristics of many machines.

5.2.1 Single Machine, Multiple Parameter Analysis

The purpose of a single machine, multiple parameter analysis on a visual basis was to evaluate generalized equations describing the relationships between machine parameters in relation to other parameters that describe the characteristics of the field and actual operation of the machines in agricultural operations. In this process, multiple machine parameters were mapped for a field with points that vary in color based on the value of the specific parameter being evaluated. Figure 13a and Figure 13b show the type of observations that can be made through the visual analysis. These Figures show the operation of a JD 8345 tractor while performing spring planting operations in an irrigated field on Farm 2. Figure 13a shows the change in the variable "Navigation" while planting. Figure 13b shows the speed variation for the same machine, field and operation. The definitions and units for Figures 13a and 13b were not properly documented in the data collection stage. The variable "Navigation" most likely relates to a machine guidance variable. The variable "Ground Speed" was likely the speed of the machine in relation to the ground. The method in which this variable was determined, however, was not known. This speed may relate to GPS speed, wheel speed, or a variety of other methods to calculate speed. The units for this variable were assumed to be "miles per hour." This assumption was based on a claim from the farm manager that the target planting speed was 5 mph (approximately 8 kph).



Figures 13a (Left) and 13b (Right) – The values of variable "Navigation" and "Ground Speed" of a JD 8345 in spring planting in an irrigated field from Farm 2: Single-machine, multiple-parameter analysis

Figures 13a and 13b show an incredible similarity between the two parameters that are plotted. The blue points in each Figure enhance the visual appearance patterns that were observed within the field. In the ground speed plot in Figure 13b, these blue points correspond to all points slower than the desired planting speed for this farm, 8 +/- 0.8 kph (5 +/- 0.5 mph). Figure 13a shows the values for "Navigation" that range from approximately 0 to 17.5+ in a blue to green scale. These plots indicate that the parameters "Navigation" and groundspeed are not only related to each other, but they are also related to the tracks of the irrigation system. An aerial view of the field without any data plots is presented in Figure 14. This Figure also contains a circular graphic concentric with the irrigation system to enhance the visual appearance of the tracks for the irrigation system and their correlation with the parameters in Figures 13a and 13b.

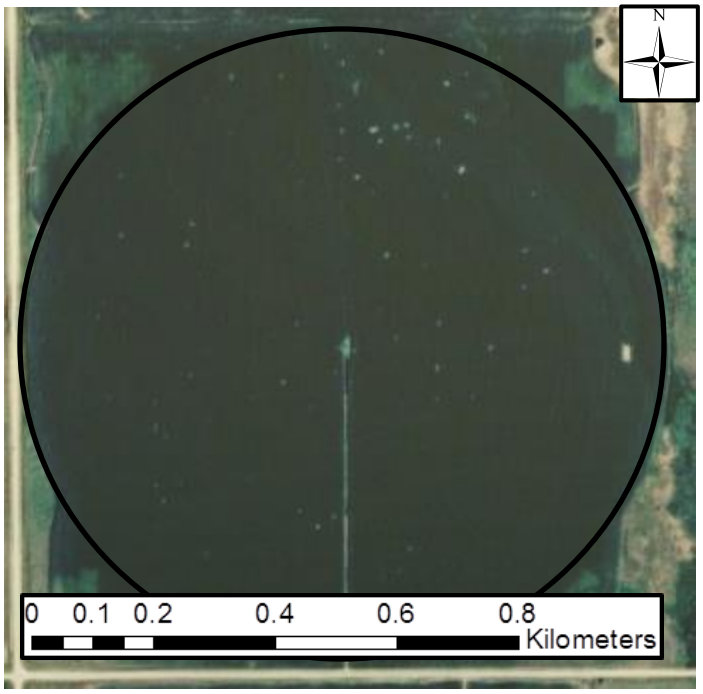
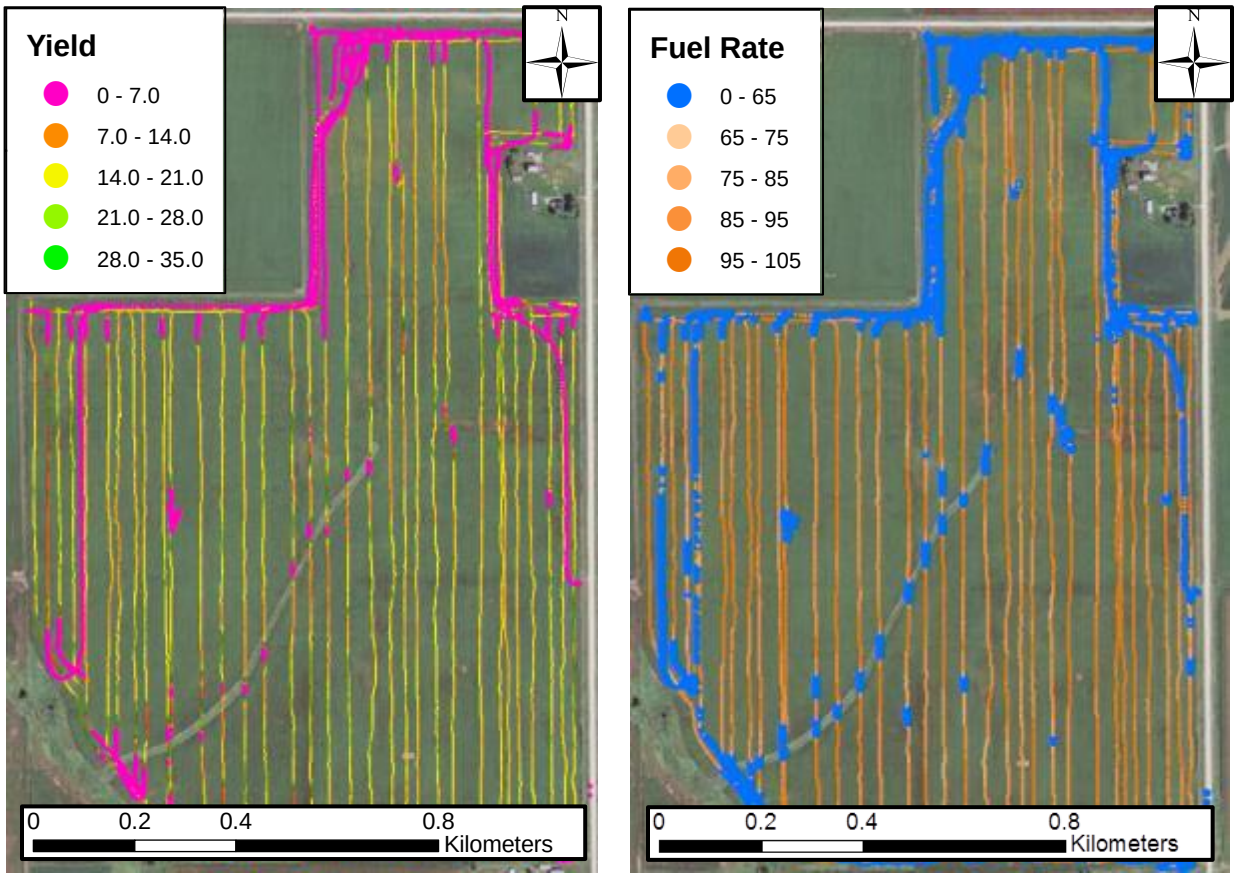


Figure 14 – Aerial view of irrigated field from Farm 2 from Figures 12a and 12b without machine data points

The ability of variable color mapping of parameters was also found useful in interpreting the data from Farm 1. The data in Figures 15a and 15b display the change in the fuel consumption and the yield respectively. The units for these variables along with the method through which their values were determined were not well documented in the data collection phase. The data indicate that the areas of low fuel consumption and low yield, which are represented by the pink points and blue points respectively, were in similar locations. The points low in magnitude appear almost solely around the perimeter of the field where the combine has traveled multiple passes, and these points can be attributed to secondary passes through the headlands where the crops had already been harvested. These low magnitude points due to their larger size and brighter color overshadow the other points in these areas. The locations of low points that are truly indicative of the correlation between these two variables are present along the waterway that leads from the middle of the field to the southwest side of the field. There are

also corresponding small groups of low points of each parameter located in the west-central, the north-central, and the east-central portions of the field.



Figures 15a (Left) and 15b (Right) – Visual inspection of low fuel consumption and low yield corresponding to pink and blue points respectively collected from a JD S680 combine in a field from Farm 1: Single-machine, multiple-parameter analysis (Ground Speed and Yield units undocumented)

5.2.2 Multiple Machine, Single Parameter Analysis

The visual evaluation of multiple parameters for one machine can be useful in influencing machine design factors and in influencing the operation of one machine. This study also includes an evaluation of the ability of ArcGIS™ to be used to describe the operation of agricultural machinery and to provide recommendations based on those descriptions to improve the efficiency or productivity of those operations. For this reason, the performance of multiple

machines within the same field must also be considered. In this study, the number of machines with detailed records of machine operating parameters was very small. The majority of machines had records of position, elevation, time, and ground speed only. This fact reduced the ability to compare the influence of terrain on a wide variety of machine parameters. In Figures 16a, 16b, and 17, the variations of planting speeds of two JD 8345 tractors in a terraced field on Farm 2 were plotted. Figures 16a and 16b display the differing paths of the two planters with speed variations ranging from 6.5 – 11.3 kph (4 – 7 mph) in the field. Green points indicate low planting speeds and red points indicate high planting speeds. The units for Figures 16a, 16b, and 17 follow the same assumption as made for Figures 13a and 13b that the target planting speed was 5 mph on this farm.

Figure 16a (Left) and 16b (Right) – Planting speed variation for two JD 8345 tractors in a terraced field on Farm 2 with green to red points corresponding to 6.5 – 11.3 kph (4 – 7 mph)

These two plots, when combined as shown in Figure 17, provide a visual basis on which the planting performance of two different machines and operators were compared. The patterns in the colors of the data points for one tractor are almost indistinguishable from those of the other tractor. With planter speed being an important factor in the quality of planting, the overall planting quality of this field should not be dependent on the planter. Instead, this plot indicates that the terraces located within this field may have a bigger impact on planting quality. In the southeastern portion of the field and the north-central portion of the field, the shape of the terraces clearly had an impact on the planter speed. The planter speed increased as the planters were going down and over the terraces whereas the planter speed decreased as the planters were going up and over the terraces.

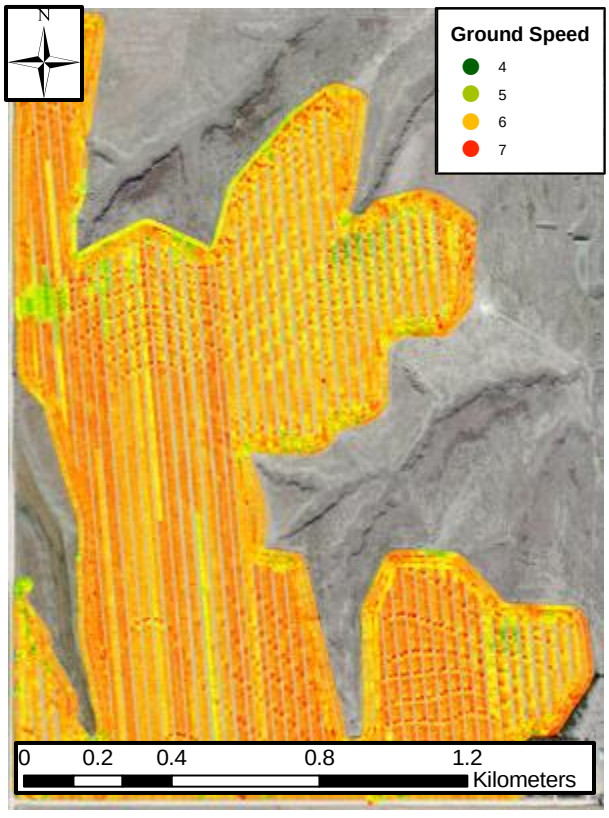


Figure 17 – Composite view of the planter speeds for two planters within a terraced field on Farm 2 as shown in Figures 16a and 16b

5.2.3 Machine Operational State Analysis

In order to evaluate the operational practices of the farms in this study, the machines were plotted as a function of three states: idle, in-field operations, and transport. This type of information provides farm managers with information regarding the area in which the machines were frequently in an idle state. It also provides engineers with a deeper understanding of the use of machines in agriculture. Design improvements may involve changing the machine so that operators are more likely to shut down the machine while not being used or improving the efficiency of the machines in these states. Figure 18 displays the changes in machine state of a JD 9770 on a small portion of Farm 2. The blue data points indicate that the machine was traveling at a transport speed greater than 13.7 kph (8.5 mph). The green data points indicate speeds at which the machine was typically in a harvest state between 3.2 – 13.7 kph (2 and 8.5 mph). Red data points indicate locations where the machine was idling or traveling at a speed slower than expected while harvesting.

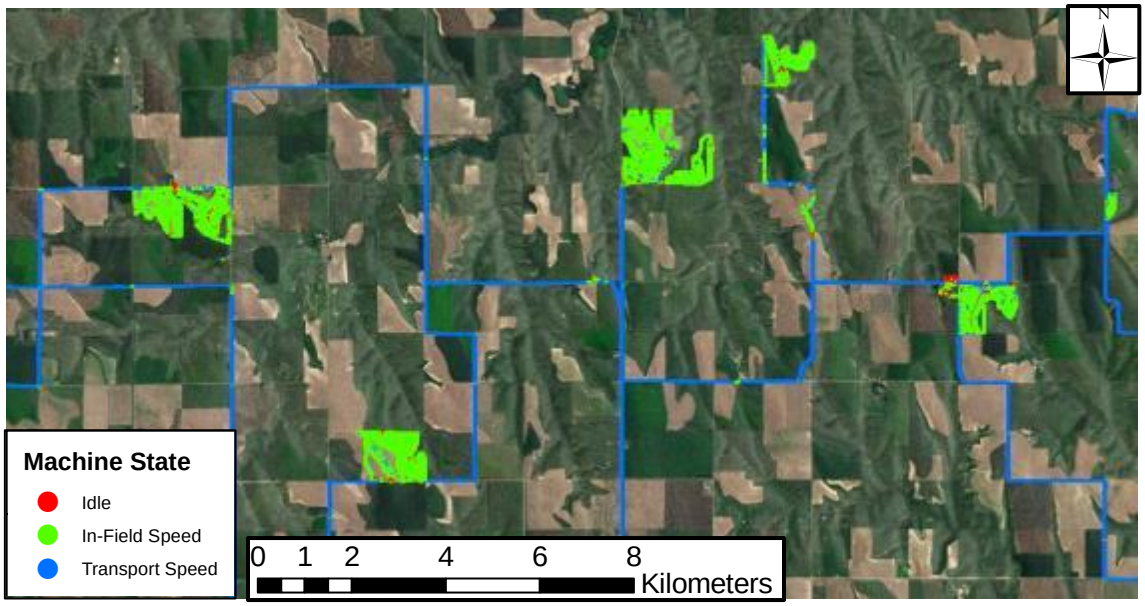


Figure 18 – Machine state for JD 9770 for a portion of operations on Farm 2

Inspection of Figure 18 indicates that there was a location in the eastern portion of this excerpt of the farm where this JD 9770 idled for a large amount of time. When inspected closer in Figure 19, the aerial imagery indicates that this concentration of idle points was a center of operations for Farm 2. If these points were determined by the farm manager to be places where there was no need for the machine to be operating, recommendations could be made to machine operators to shut the machine down while in this location to prevent unnecessary fuel consumption and machine hours from accumulating on the machine.

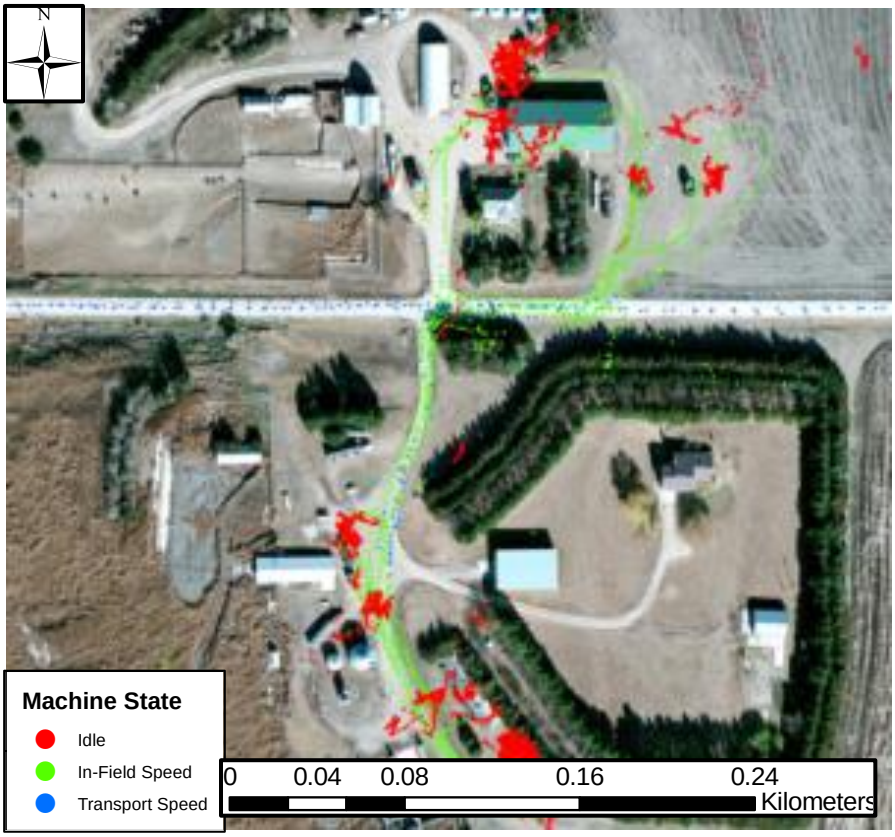


Figure 19 – Close view of idling points of JD 9770 depicted in the east portion of Figure 18

5.2.4 Multiple Machine Interaction Analysis

The interaction between multiple machines can also provide engineers and farm managers with useful information regarding the operation of machines. Specifically of interest

was the interaction between the combines and tractors with grain carts while unloading on the go. From information regarding unloading on the go, farm managers could develop strategies for minimizing down time of combines while waiting to unload by identifying the average time for an empty grain cart to travel to the combine, for the combine to unload into the grain cart, for the full grain cart to travel back to the semi-truck, and then for the grain to be unloaded into the semi-truck from the grain cart.

As seen in Figure 20, the stages of the process are easily identified using the variable color mapping in ArcGIS™. The black and pink points indicate the path that the JD S680 combine took while harvesting this field on Farm 1. The green to yellow to red points indicate the path that the JD 8360 tractor with a grain cart took. These points indicate the speed of the tractor from green at the slowest points to red and the fastest points. The black path points for the combine indicate locations where the combine's unloading auger was not engaged. The pink points indicate the locations where the unloading auger was engaged and the combine was unloading grain into the grain cart. One instance of the combine unloading while still harvesting in the field is specified by the black oval in Figure 20. In this location of the field, the tractor positioned the grain cart parallel to the combine's path in order for the combine to unload grain into the grain cart. The reason for the lack of corresponding tractor and grain cart paths for every occurrence of the combine unloading was due to the lack of data quality. There were other combines and tractors operating within this field on this date that did not collect sufficient data with enough quality to display their unloading characteristics.

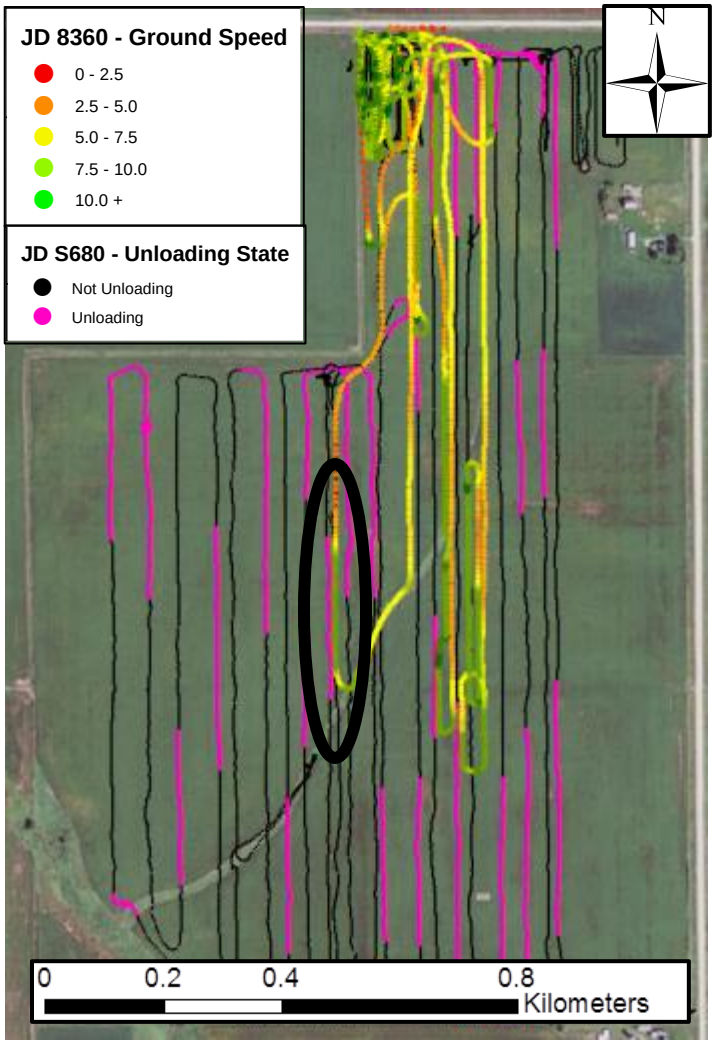


Figure 20 – JD S680 combine and JD 8360 tractor with grain cart in harvesting operations on Farm 1: Black ellipse showing one instance of the combine shown unloading into the tractor and grain cart shown (Ground Speed units undocumented)

5.3 Statistical Analysis

The observation of trends within the data on a visual basis was an important step to identify unexpected relationships between parameters that may not be readily apparent through other forms of inspecting the data. This method of describing machine parameters did not properly describe the data in a quantitative measure. In order to more effectively describe the data quantitatively, three statistical analyses were performed on the data. First, the visual inspection of the data was supported with the application of linear regression models to

determine the validity of relations observed. The application of linear regression models to only the observed correlations would have assumed that all correlations had been visually identified. To identify the correlations that may not have been observed, a second method of statistical analysis on the data was performed. This method consisted of viewing the mean and standard deviation of single machine parameters on a field-by-field basis for other influencing factors on machine performance. Finally, the statistics regarding the operational state of the machine were generated to provide additional information about the use of machines in farming operations. The combination of the metrics generated through these methods allows engineers to alter designs and farm managers to alter operation strategies in order to reduce costs, increase efficiency, increase productivity, and otherwise improve equipment and operations.

5.3.1 Single Machine, Multiple Parameter Linear Regression Analysis

Linear regression models were created for parameters that were considered influential in the operation of the machines. Within a data set for a single machine, the parameters of interest typically related to productivity, cost, or efficiency of the machine. In order to provide a complete analysis on the correlations between parameters, the data set describing a particular machine's operation must contain a variety of independent variables that would be expected to influence the dependent variable. Fuel consumption was of particular interest in this study, but the lack of data quality and incomplete records of relevant parameters from the machine impeded the ability to use linear regression to properly analyze the influences on fuel consumption.

The following example of linear regression analysis of the data was provided in order to illustrate the capability of ArcGIS™ to generate linear regression models. Conclusions based on these data other than the capability of ArcGIS™ to assist in making decisions were not intended to accurately describe the operating characteristics of the machines. One of the machines with

the greatest amount of parameters that were recorded was a JD 9770 combine on Farm 2. The objective of the linear regression analysis of this machine's data was to determine the factors that influence the rate at which fuel was consumed. Seven variables were considered in this analysis: unloading auger engagement, engine speed, header position, machine pitch, ground speed, unloading auger speed, and yield. Not all of these parameters were expected to influence the fuel consumption, and the confirmation that they were not correlated with fuel consumption served as proof that the linear regression model would confirm and deny correlations.

The results of single variable linear regression analyses are presented in Table 3. This table displays the coefficient of determination for the single variable linear regression models for the seven variables shown along with a multiple variable linear regression model that included all variables tested in a single variable method. The output of this model indicated that engine speed and machine speed were the variables that were most strongly related to the fuel consumption rate explaining 62% and 35% of the variation in the fuel consumption respectively. The multiple variable analysis included the three variables that showed the most significant effect on fuel consumption: engine speed, ground speed, and machine pitch. The intent of the multiple variable analysis was to indicate the ability of the statistical tools within ArcGIS™ to evaluate the effect of known machine parameters as well as environmental variables on the fuel consumption rate. Example output files from which the data in Table 3 originated are included in Appendix C. These output files include other statistical descriptors of the models, graphs of the possible explanatory variables versus the dependent variable fuel consumption rate, graphs of residual distributions, graphs of residuals versus the values predicted by the model, and the input parameters for the models. A coefficient of determination of 1 would have indicated that there was a direct linear relationship between the two parameters.

Table 3 – Coefficient of determinations for the regression analyses of fuel consumption rate for a JD 9770 combine on Farm 2 as affected by selected variables

Fuel Consumption Rate Linear Regression Analysis	
JD 9770 - Farm 2	
Single Variable Analysis	Coefficient of Determination
Auger Engagement	0.01190
Engine Speed	0.62235
Header Position	0.00410
Machine Pitch	0.09229
Machine Speed	0.35039
Unloading Auger Speed	0.07048
Yield	0.00061
Multiple Variable Analysis	0.70026

The variables indicated in Table 3 were not well documented in the collection of the parameter values. Observation of changes in the data as well as inferences that could be drawn from the names of the variables allowed a general conclusion to be made of what each parameter indicated. The "Auger Engagement" was assumed to describe an on or off value (1 or 0) for the engagement of the unloading auger on the combine. "Engine Speed" was assumed to be the speed of the combine's engine with units most likely being rpm. The variable "Header Position" likely described the position of the header in an up or down value also associated with either a 1 or 0 value, respectively. The "Machine Pitch" was determined to be the pitch of the machine in some arbitrary scale. The "Machine Speed" likely indicated the actual ground speed of the machine measured by unknown means. The "Unloading Auger Speed" appeared to be the percentage of maximum rotational speed of the combine's unloading auger with a value "0" corresponding to a value of 0 for the "Auger Engagement" variable, and a value of 95-100 associated with the value 1 for "Auger Engagement." The variable "Yield" was assumed to be the instantaneous yield of the crop as measured by the on-board yield monitor.

5.3.2 Single Machine, Multiple Field Performance Metrics

The use of single and multiple variable linear regression analyses provided useful examples of how the relationships between machine parameters could be determined. These relationships could be useful in improving the overall operation of the machine, but they implied nothing regarding the performance of the machine across multiple fields. In order to determine the performance of a machine in different fields, a comparison of operational statistics was performed across multiple fields. The comparison of the mean and standard deviation of parameters recorded in individual fields to those recorded on the entire farm allowed the influence of field shape, logistics, operator, and other non-numeric factors to be included in the determination of machine performance.

The statistics for the ground speed for the JD 4830 sprayer on Farm 2 are shown for a portion of the fields in Table 4 and in Figure 21. The mean ground speed for all fields as indicated by both Table 4 and Figure 21 was higher than all of the individual fields due to the fact that this number includes transport speeds between fields. Additionally, the standard deviation was expected to be higher for the entire farm than for the individual fields due to the large difference between the high speeds occurring during transport and the zero speeds that occur while the machine was idling.

In order to increase the consistency of productivity with which field operations occur across fields, an attempt should be made to reduce the standard deviation of the ground speed. This type of data presentation could be useful for farm managers to increase the consistency of operations across all fields by identifying the fields for which an operating parameter, ground speed in this case, is much different than normal.

Table 4 – JD 4830 sprayer ground speed statistics for Farm 2

Farm 2 - JD 4830 Sprayer		
Ground Speed Statistics (Units Undocumented)		
Field	Mean	Standard Dev.
All Fields	11.5	7.0
A	9.4	5.9
B	5.9	5.5
C	7.1	6.9
D	7.1	5.8
E	8.7	6.2
F	10.9	5.5
G	8.7	6.5
H	7.2	4.8
I	7.8	6.1
J	9.1	5.9

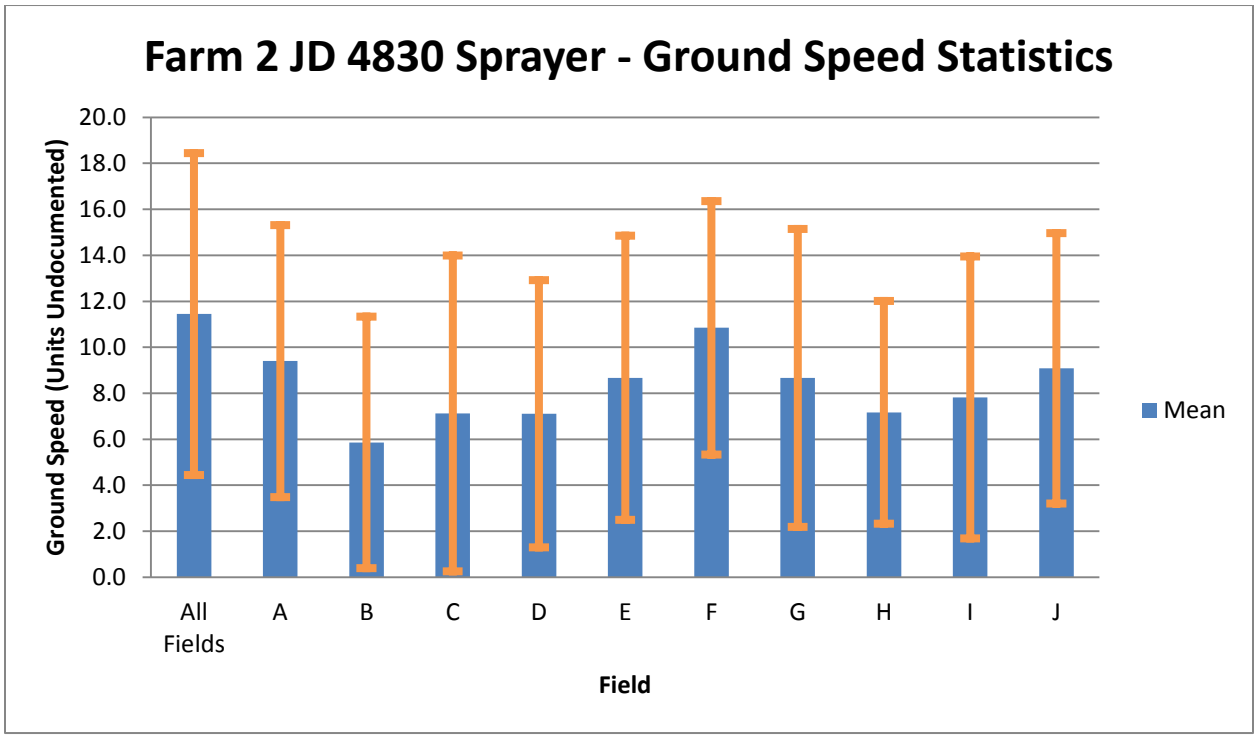


Figure 21 – JD 4830 mean ground speeds by field with error bars including one standard deviation above and below mean

The statistics shown in Table 4 and Figure 21 indicated that field F showed the highest rate of productivity with a low standard deviation of the ground speed. In order to improve consistency across fields, the attempt should be made to raise the mean ground speed of the sprayer in other fields to the value for field F. The results of analyzing these ten fields in combination with the farm as a whole also indicated that field B showed the lowest rate of productivity, and it also showed a high standard deviation of speeds compared to that average. Visual inspection of field F (Figure 22) and field B (Figure 23) provided explanations of the reasons for the large difference in productivity rates and standard deviation for these two fields.

Field F was a square shaped field. The geometry of this field allowed the sprayer to reduce its speed minimally in order to turn in the headlands. The lack of significant changes in elevation also allowed the sprayer to maintain a steady high speed within the field. The shape of field B also provided an explanation of the reason for the groundspeed of the sprayer. Field B was composed of the four corners outside an irrigated field. The corners of this field required different spraying applications than the central irrigated portion due to two factors. The corners are often planted with a different crop than the irrigated portion of the field because of the water requirements of different crops. In the case that the corners of the fields are planted in the same crop as the irrigated portion, chemical application differs due to the requirements of crops with less water than those that are irrigated. The separation of the corners from the central portion of the field required the sprayer to make more turns within the field compared to the time spent in the central portion reducing the mean speed and increasing the variability in the speed.

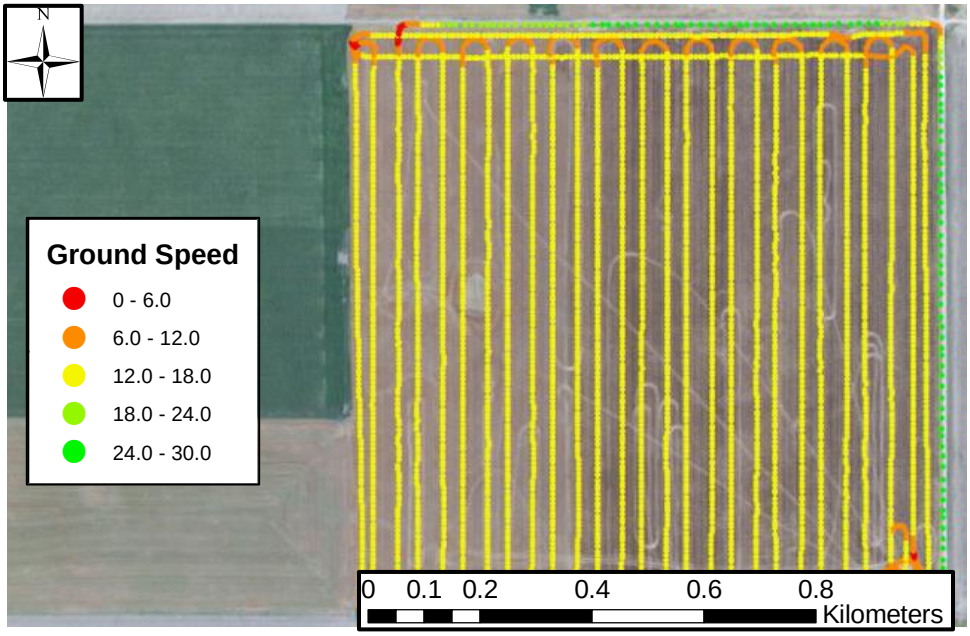


Figure 22 – JD 4830 ground speed in field F on Farm 2 (Ground Speed units undocumented)

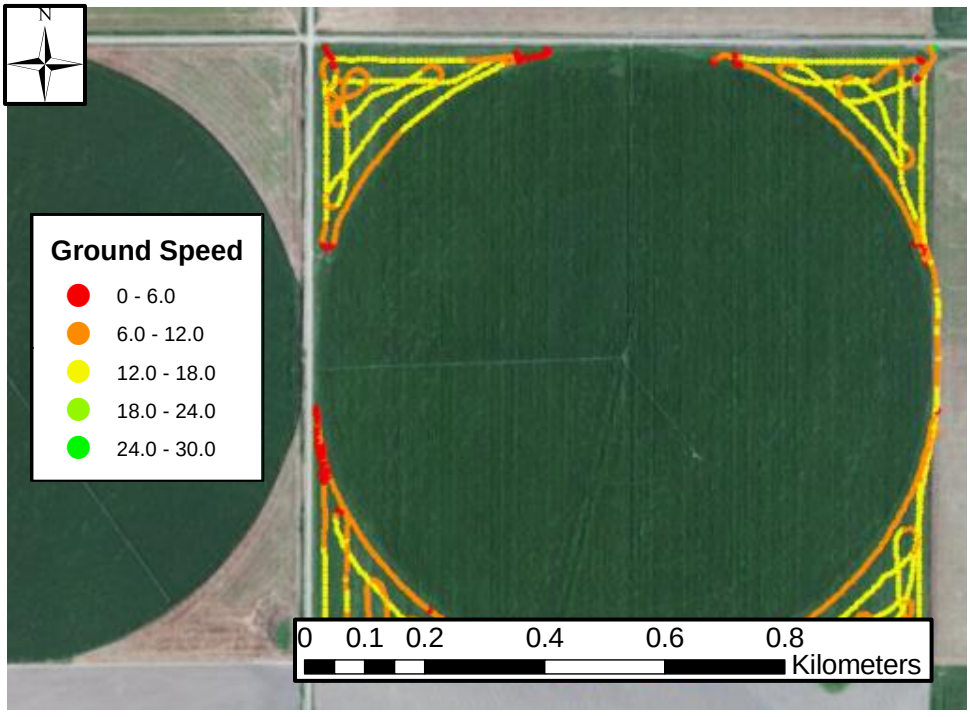


Figure 23 – JD 4830 ground speed in field B on Farm 2 (Ground Speed units undocumented)

The conclusion from these results did not take into account the change in quality in spraying in recommending an increase in speed. It also did not imply that the geometry of the field should be changed in order to facilitate a higher productivity rate. These results were intended to display the effects of varied field characteristics on the machine parameters. The application of this analysis with higher quality data in the future is intended to provide a recommendation to farmers as to the impact of field characteristics on production rates and costs.

5.3.3 Multiple Machine Operation State Metrics

The results in previous sections discuss the operating characteristics of machines themselves, but they neglected to account for the characteristics of the people that operate these machines. One of the largest areas of machine inefficiency was determined to be the idle state where the machine is being operated, but it was not productive in accomplishing any tasks. In order to assess the location of these effects, the machine states were displayed in a visual manner in Figures 22 and 23. To assess the magnitude of the effects of the idle states, statistical descriptions regarding the idle, in-field, and transport machine states were generated. In this study, the state of the machine was determined on a speed related basis. The idle state was defined to be the speeds less than 3.2 kph (2 mph) where the machine was considered to be either stationary or unproductive. The in-field operation speeds were defined through the analysis of speeds at which the machines operated in field. The transport speeds of the machines were defined as any speed greater than those expected of in-field operation. The definitions for idle, in-field, and transport states are given for Farm 1 and Farm 2 in Tables 5a and 5b.

Table 5a – Machine state definitions for operations on Farm 1

State Speeds, kph (mph)			
Machine	Idle	In-Field	Transport
S680 - 1	0 - 3.2 (0 - 2.0)	3.2 - 13.7 (2.0 - 8.5)	13.7 + (8.5 +)
S680 - 2	0 - 3.2 (0 - 2.0)	3.2 - 13.7 (2.0 - 8.5)	13.7 + (8.5 +)
9870	0 - 3.2 (0 - 2.0)	3.2 - 13.7 (2.0 - 8.5)	13.7 + (8.5 +)
9460 - 1	0 - 3.2 (0 - 2.0)	3.2 - 12.9 (2.0 - 8.0)	12.9 + (8.0 +)
9460 - 2	0 - 3.2 (0 - 2.0)	3.2 - 12.9 (2.0 - 8.0)	12.9 + (8.0 +)
9460 - 3	0 - 3.2 (0 - 2.0)	3.2 - 12.9 (2.0 - 8.0)	12.9 + (8.0 +)
9460 - 4	0 - 3.2 (0 - 2.0)	3.2 - 12.9 (2.0 - 8.0)	12.9 + (8.0 +)
9460 - 5	0 - 3.2 (0 - 2.0)	3.2 - 12.9 (2.0 - 8.0)	12.9 + (8.0 +)
8360 - 1	0 - 3.2 (0 - 2.0)	3.2 - 14.5 (2.0 - 9.0)	14.5 + (9.0 +)
8360 - 2	0 - 3.2 (0 - 2.0)	3.2 - 14.5 (2.0 - 9.0)	14.5 + (9.0 +)
8360 - 3	0 - 3.2 (0 - 2.0)	3.2 - 14.5 (2.0 - 9.0)	14.5 + (9.0 +)
8360 - 4	0 - 3.2 (0 - 2.0)	3.2 - 14.5 (2.0 - 9.0)	14.5 + (9.0 +)
8360 - 5	0 - 3.2 (0 - 2.0)	3.2 - 14.5 (2.0 - 9.0)	14.5 + (9.0 +)
6170 - 1	0 - 3.2 (0 - 2.0)	3.2 - 14.5 (2.0 - 9.0)	14.5 + (9.0 +)
6170 - 2	0 - 3.2 (0 - 2.0)	3.2 - 14.5 (2.0 - 9.0)	14.5 + (9.0 +)

Table 5b – Machine state definitions for operations on Farm 2

State Speeds, kph (mph)			
Machine	Idle	In-Field	Transport
8345 - 1	0 - 3.2 (0 - 2.0)	3.2 - 14.5 (2.0 - 9.0)	14.5 + (9.0 +)
8345 - 2	0 - 3.2 (0 - 2.0)	3.2 - 14.5 (2.0 - 9.0)	14.5 + (9.0 +)
8345 - 3	0 - 3.2 (0 - 2.0)	3.2 - 14.5 (2.0 - 9.0)	14.5 + (9.0 +)
9770 - 1	0 - 3.2 (0 - 2.0)	3.2 - 13.7 (2.0 - 8.5)	13.7 + (8.5 +)
9770 - 2	0 - 3.2 (0 - 2.0)	3.2 - 13.7 (2.0 - 8.5)	13.7 + (8.5 +)
9770 - 3	0 - 3.2 (0 - 2.0)	3.2 - 13.7 (2.0 - 8.5)	13.7 + (8.5 +)
4830	0 - 3.2 (0 - 2.0)	3.2 - 24.1 (2.0 - 15.0)	24.1 + (15.0 +)

The metrics generated by applying the definitions for machine states are displayed in Tables 6a and 6b for Farms 1 and 2, respectively. These metrics indicated that the machines were operated a significant amount of time in the idle state. Some machines spent less than 50% of their total hours actually operating in-field. These values enable farm managers to identify operators that are idling their machines for a longer amount of time than the farm average and to

assist them in reducing the amount of unnecessary fuel burned along with the machine hours acquired during these states. A breakdown of the average percent of time in each machine state for Farm 1 and Farm 2 is shown in Tables 7a and 7b. Also included are the minimum and maximum percentages of time spent in each state and the standard deviation of each state. The machines did not spend the same total amount of time operating, and the numbers indicated in Tables 7a and 7b do not compensate for this.

Table 6a – Machine state breakdown for Farm 1 according to rules in Table 5a

Machine	Time in State (hours)				Percent Time in State		
	Idle	In-Field	Transport	All	Idle	In-Field	Transport
S680 - 1	41.4	209.7	20.6	271.7	15%	77%	8%
S680 - 2	82.3	96.1	7.3	185.7	44%	52%	4%
9870	10.1	44.4	10.3	64.9	16%	68%	16%
9460 - 1	30.6	147.6	19.2	197.4	15%	75%	10%
9460 - 2	20.6	111.8	16.8	149.2	14%	75%	11%
9460 - 3	26.3	171.2	33.5	231.0	11%	74%	15%
9460 - 4	13.7	90.3	10.4	114.3	12%	79%	9%
9460 - 5	44.5	332.3	41.2	418.0	11%	79%	10%
8360 - 1	169.8	190.4	57.8	418.0	41%	46%	14%
8360 - 2	177.4	195.5	114.4	487.4	36%	40%	23%
8360 - 3	47.2	105.8	42.9	195.9	24%	54%	22%
8360 - 4	158.1	221.7	60.6	440.5	36%	50%	14%
8360 - 5	41.3	71.7	22.1	135.0	31%	53%	16%
6170 - 1	24.4	42.9	7.1	74.4	33%	58%	9%
6170 - 2	21.3	43.1	9.3	73.7	29%	59%	13%

Table 6b – Machine state breakdown for Farm 2 according to rules in Table 5b

Machine	Time in State (hours)				Percent Time in State		
	Idle	In-Field	Transport	All	Idle	In-Field	Transport
8345 - 1	17.1	60.3	18.5	95.9	18%	63%	19%
8345 - 2	73.6	194.2	23.1	290.9	25%	67%	8%
8345 - 3	123.3	167.2	47.9	338.5	36%	49%	14%
9770 - 1	34.6	108.2	19.7	162.5	21%	67%	12%
9770 - 2	39.5	129.3	37.5	206.3	19%	63%	18%
9770 - 3	80.3	278.3	88.6	447.2	18%	62%	20%
4830	68.9	206.9	24.0	299.8	23%	69%	8%

Tables 7a (Left) and 7b (Right) – Machine state statistics for results shown in Tables 6a and 6b for Farm 1 and Farm 2, respectively

All Machines, Farm 1 - State Statistics			
	Idle	In-Field	Transport
Mean	25%	63%	13%
Std. Dev.	12%	13%	5%
Min.	11%	40%	4%
Max.	44%	80%	24%

All Machines, Farm 2 - State Statistics			
	Idle	In-Field	Transport
Mean	23%	63%	14%
Std. Dev.	7%	6%	5%
Min.	18%	49%	8%
Max.	36%	69%	20%

CHAPTER 6: SUMMARY AND CONCLUSIONS

Field performance of agricultural machinery was recorded from tractors, sprayers, and combines from two Midwestern farms. The data were collected with customized versions of the John Deere StarFire™ and GreenStar™ precision agriculture solution and a customized version of the MTG Data Logger used in the JDLink™ system. The machine performance was collected in order to provide one year's worth of performance data describing the operations of agricultural machinery to researchers seeking to improve the overall efficiency of agricultural operations.

After collecting the data from the machinery, John Deere distributed the data to the researchers involved in this study. When received, the data were preprocessed to assess the quality of the data and to prepare the data for inspection according to the objectives of improving agricultural operations efficiency. After preprocessing, the data were imported into Esri's ArcGIS™ to perform a geospatial analysis of the data quality and the performance of the agricultural machinery.

After assessing the quality of the data collected in this study, it was determined that the data collected in this study was lacking in completeness. This assessment was made based on the preprocessing stage of analysis of the data. This stage determined that approximately 44% of data on Farm 1 and 37% of data on Farm 2 were improperly collected, recorded, transmitted, or maintained. As a result, this study focused on a method with which future data could be interpreted instead of the recommendation of altered machine designs or operational procedures.

Analysis of the data that were collected indicated that the machines on Farm 1 averaged 25% of operating hours in the idle state and 63% and 13% in the in-field and transport states respectively. The data for Farm 2 indicated similar percentages for each state with idle, in-field,

and transport times representing 23.0%, 62.8%, and 14.2% averaged over all machines and fields. In future years of John Deere's study of agricultural operations efficiency, it is anticipated that the data quality will be increased drastically with the implementation of recommendations presented in this study. The statistical analysis of the data was also able to explain almost 72% of variation of fuel consumption of a combine by the variation of other machine parameters in relation to the fuel consumption rate. With improved data, the method presented in this study can be implemented to provide a much better qualitative and quantitative assessment of machine performance in agricultural operations. From that assessment, design improvements for agricultural machinery will be able to be made, and operations strategies will be created that increase operational efficiency. The conclusions of this study according to each objective are as follows:

- Identify anomalies in machine performance – The ability to detect anomalies in machine performance in this study was not able to be accomplished due to incomplete records of machine performance data and the incomplete documentation of machine parameter definitions.
- Determine relationships between machine parameters – The relationships between machine parameters were not able to be determined due to incomplete documentation of parameter definitions and units. The ability to determine the relationships between parameters was shown with the data that were available.
- Provide explanations of those connections between parameters to both John Deere and the farm managers from which the data were collected – Explanations regarding the relationships between parameters could not be provided based on the lack of parameter definitions.

- Develop both design and operational recommendations to John Deere and the farm managers from which the data were collected – Recommendations to improve both designs of machines and the operations on the farms were not made due the low quality of the data and the low confidence of resulting conclusions.
- Compare operations in terms of performance metrics – The ability to compare the performance of machinery on both farms was accomplished by machine state only. The comparison of the percentage of time spent in each machine state for Farm 1 and Farm 2 showed that the machine utilization on each farm was very close. Additional detail regarding the comparison between the two farms was not possible due to the low number of farms compared.

CHAPTER 7: RECOMMENDATIONS

The evaluation of John Deere's data collection method and the method developed in this study to analyze that data resulted in several recommendations. The following points and explanations should be considered in future research in this area:

1. On-machine data storage – One of the major complications of this study was the poor quality of the data. It is anticipated that storing a copy of the data on the machine either in addition or in replacement of the cellular transmission of the data will provide a more complete set of data for analysis.
2. Additional machine parameter collection – In the first season's collection of machine operating parameters, the data collected were mostly limited to the position, speed, and time for each data point. Collecting additional machine parameters relevant to improving machine efficiency will allow more in-depth conclusions to be made upon the data. The selection of parameters should be made based on the specific parameters able to be recorded from the machines and the changing objectives of the overall John Deere study.
3. Machine parameter standardization – The majority of the tasks related to preprocessing the data, generating maps, and providing statistical analysis of the data could have been simple to analyze. The number and order of the parameters collected for each type of machine were changed multiple times for each machine throughout the season. In order to automate the data analysis process, the setup for each model of machine should be standardized.
4. StarFire™ and GreenStar™ data collection – The data collected via the MTG data logger was associated with file types of significantly less quality than the data collected with the StarFire™ and GreenStar™ system. The collection of geospatial position with the

StarFire™ receiver provides increased accuracy over the position recorded with the MTG Data Logger, and the GreenStar™ data display increased the ease with which the data could be imported into ArcGIS™.

5. Independent machine operation logs – The lack of definitive records associated with the operation of the machines in field operations decreased the ability to compare the computer collected data to reality. The collection of operator, fuel use, operation time, and other metrics would provide an additional basis on which the data quality can be compared. This information would also assist in identifying operators and fields that are associated with more efficient operations.
6. Documentation of parameters – Some machine parameters recorded in the data collection process were very vaguely described. The name of the parameter and the units of the parameter were recorded in the setup files for the data collection devices, but a detailed description of these parameters was not included. In order to provide reliable recommendations based on the data, the parameters must be documented more clearly in the future to avoid misinterpretation of the parameters or confusion enhanced by multiple machines having parameters with the same name but different descriptions.
7. Repeat of data analysis method – After observing the first season's data for this study, it was determined that the method included in this study was the optimal strategy for accomplishing the objectives included in Chapter 2. The repetition of this method with subsequent years' data collection will either validate its soundness or provide additional recommendations through which the method can be improved.
8. Application of new machine state rules – With the implementation of recommendation number two in this section, new definitions for machine state can be implemented in

ArcGIS™ in order to more accurately define the state of each machine. The recommended definitions for machine state are included in Appendix D.

9. Record wireless communication metrics – In order to determine the cause of data quality loss, metrics pertinent to the wireless communication of data including cellular signal quality and transmission speed should be collected.

With the implementation of these recommendations, it is anticipated that the evaluation of machine performance will be dramatically improved. The quality of the data, the conclusions based upon the data, and the resulting recommendations will improve future research into the area of agricultural machine performance evaluation.

REFERENCES

Alkobaisi, S., W. Bae, P. Vojtěchovský and S. Narayanappa. 2012. An Interactive Framework for Spatial Joins: A Statistical Approach to Data Analysis in GIS. *GeoInformatica* 16(2): 329-355.

ASABE D497.7. 2011. Agricultural Machinery Management Data. *ASABE* D497.7.

Boon, N. E., A. Yahya, A. F. Kheiralla, B. S. Wee and S. K. Gew. 2005. A Tractor-Mounted, Automated Soil Penetrometer–Shearometer Unit for Mapping Soil Mechanical Properties. *Biosystems Engineering* 90(4): 381-396.

Chang, Y. K., Q. Zaman, A. A. Farooque, A. W. Schumann and D. C. Percival. 2012. An Automated Yield Monitoring System II for Commercial Wild Blueberry Double-Head Harvester. *Computers and Electronics in Agriculture* 81(0): 97-103.

Chaudhry, O. Z. and W. A. Mackaness. 2010. DTM Generalisation: Handling Large Volumes of Data for Multi-Scale Mapping. *Cartographic Journal* 47(4): 360-370.

Chen, Y., J. Yu and S. Khan. 2010. Spatial Sensitivity Analysis of Multi-Criteria Weights in GIS-Based Land Suitability Evaluation. *Environmental Modelling & Software* 25(12): 1582-1591.

Clemmer, G. 2010. *The GIS 20: Essential Skills.* Redlands, California: Esri Press.

Devillers, R., Y. Bédard, R. Jeansoulin and B. Moulin. 2007. Towards Spatial Data Quality Information Analysis Tools for Experts Assessing the Fitness for Use of Spatial Data. *International Journal of Geographical Information Science* 21(3): 261-282.

Duttmann, R., J. Brunotte and M. Bach. 2013. Spatial Analyses of Field Traffic Intensity and Modeling of Changes in Wheel Load and Ground Contact Pressure in Individual Fields During a Silage Maize Harvest. *Soil and Tillage Research* 126(0): 100-111.

Farooque, A. A., Y. K. Chang, Q. U. Zaman, D. Groulx, A. W. Schumann and T. J. Esau. 2013. Performance Evaluation of Multiple Ground Based Sensors Mounted on a Commercial Wild Blueberry Harvester to Sense Plant Height, Fruit Yield and Topographic Features in Real-Time. *Computers and Electronics in Agriculture* 91(0): 135-144.

Goering, C. and A. Hansen. 2008. *Engine and Tractor Power.* St. Joseph, MI: American Society of Agricultural and Biological Engineers.

Goering, C., M. Stone, D. Smith and P. Turnquist. 2006. *Off-Road Vehicle Engineering Principles.* St. Joseph, MI: American Society of Agricultural Engineers.

Gorr, W. and K. Kristen. 2011. *GIS Tutorial 1: Basic Workbook.* Redlands, California: Esri Press.

Grogan, J., D. A. Morris, S. W. Searcy and B. A. Stout. 1987. Microcomputer-Based Tractor Performance Monitoring and Optimization System. *Journal of Agricultural Engineering Research* 38(4): 227-243.

Hao, M., D. A. Keim, U. Dayal, D. Oelke and C. Tremblay. 2008. Density Displays for Data Stream Monitoring. *Computer Graphics Forum* 27(3): 895-902.

Hawick, K. A., P. D. Coddington and H. A. James. 2003. Distributed Frameworks and Parallel Algorithms for Processing Large-Scale Geographic Data. *Parallel Computing* 29(10): 1297-1333.

Jacobs, A. 2009. The Pathologies of Big Data. *Communications of the ACM* 52(8): 36-44.

Jasiewicz, J. 2011. A New GRASS GIS Fuzzy Inference System for Massive Data Analysis. *Computers & Geosciences* 37(9): 1525-1531.

John Deere. 2013a. GreenStar (TM) 3 Display. Available at: http://www.deere.com/wps/dcom/en_US/products/equipment/ag_management_solutions/displays_and_receivers/greenstar_3_display_2630/greenstar_3_display_2630.page?. Accessed March 5 2013.

John Deere. 2013b. StarFire (TM) 3000. Available at: http://www.deere.com/wps/dcom/en_US/products/equipment/ag_management_solutions/displays_and_receivers/starfire_3000/starfire_3000.page?. Accessed March 5 2013.

John Deere. 2013c. JDLink (TM) Machine Monitoring. Available at: http://www.deere.com/wps/dcom/en_US/services_and_support/product_support/construction_technology_solutions/machine_monitoring/machine_monitoring.page. Accessed March 5 2013.

John Deere. 2013d. JDLink (TM). Available at: http://www.deere.com/wps/dcom/en_US/products/equipment/ag_management_solutions/information_management/jdlink/jdlink.page. Accessed March 5 2013.

Kimmance, J. P., M. P. Bradshaw and H. H. Seetoh. 1999. Geographical Information System (GIS) Application to Construction and Geotechnical Data Management on MRT Construction Projects in Singapore. *Tunnelling and Underground Space Technology* 14(4): 469-479.

Kumi-Boateng, B. and I. Yakubu. 2010. Assessing the Quality of Spatial Data. *European Journal of Scientific Research* 43(4): 507-515.

Laskey, K. B., E. J. Wright and P. C. G. da Costa. 2010. Envisioning Uncertainty in Geospatial Information. *International Journal of Approximate Reasoning* 51(2): 209-223.

Li, D., J. Zhang and H. Wu. 2012. Spatial Data Quality and Beyond. *International Journal of Geographical Information Science* 26(12): 2277-2290.

Liu, B. and A. Tuzhilin. 2008. Managing Large Collections of Data Mining Models. *Communications of the ACM* 51(2): 85-89.

Louisiana Department of Transportation and Development. 2008. Construction Plans Quality Control / Quality Assurance Manual. Available at: http://www.dotd.louisiana.gov/highways/project_devel/design/documents/const_plans_qc-qa_manual.pdf. Accessed April 17 2013.

Marinos, G. 2004. Enticing But Dangerous: Assessing Web Services From a Data Quality Perspective. *DM Review* 14(5): 27-29.

Matějíček, L., P. Engst and Z. Jaňour. 2006. A GIS-Based Approach to Spatio-Temporal Analysis of Environmental Pollution in Urban Areas: A Case Study of Prague's Environment Extended by LIDAR Data. *Ecological Modelling* 199(3): 261-277.

McKinion, J. M., J. L. Willers and J. N. Jenkins. 2010. Spatial Analyses to Evaluate Multi-Crop Yield Stability for a Field. *Computers and Electronics in Agriculture* 70(1): 187-198.

Mendas, A. and A. Delali. 2012. Integration of MultiCriteria Decision Analysis in GIS to Develop Land Suitability for Agriculture: Application to Durum Wheat Cultivation in the Region of Mleta in Algeria. *Computers and Electronics in Agriculture* 83(0): 117-126.

Moody, L., H. Li, R. Burns, H. Xin and R. Gates. 2006. Quality Assurance Project Plan (QAPP) for Monitoring Gaseous and Particulate Matter Emissions from Southeastern Broiler Houses. In *Air and Waste Management Association - Symposium on Air Quality Measurement: Methods and Technology 2006,* 470-481.

Nahm, M. L., C. F. Pieper and M. M. Cunningham. 2008. Quantifying Data Quality for Clinical Trials Using Electronic Data Capture. *PLoS Clinical Trials* 5(8): 1-8.

Niehaus, C. 2013. Unpublished Research Notes.

Ormsby, T., E. Napoleon, R. Burke, C. Groessl and L. Bowden. 2010. *Getting to Know ArcGIS Desktop.* Redlands, California: Esri Press.

Paradice, D. B. and W. L. Fuerst. 1991. An MIS Data Quality Methodology Based on Optimal Error Detection. *Journal of Information Systems* 5(1): 48-66.

Parssian, A., S. Sarkar and V. S. Jacob. 2004. Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product. *Management Science* 50(7): 967-982.

Pundt, H. and K. Brinkkötter-Runde. 2000. Visualization of Spatial Data for Field Based GIS. *Computers & Geosciences* 26(1): 51-56.

Resop, J. P., D. H. Fleisher, Q. Wang, D. J. Timlin and V. R. Reddy. 2012. Combining Explanatory Crop Models with Geospatial Data for Regional Analyses of Crop Yield Using Field-Scale Modeling Units. *Computers and Electronics in Agriculture* 89(0): 51-61.

Rzevski, G. 2011. A Practical Methodology for Managing Complexity. *Emergence: Complexity & Organization* 13(1): 38-56.

Schmidt, J. P., R. K. Taylor and R. J. Gehl. 2003. Developing Topographic Maps Using a Sub-Meter Accuracy Global Positioning Receiver. *Applied Engineering in Agriculture* 19(3): 291-300.

Singh, C. D. and R. C. Singh. 2011. Computerized Instrumentation System for Monitoring the Tractor Performance in the Field. *Journal of Terramechanics* 48(5): 333-338.

Srivastava, A., C. Goering, R. Rohrbach and D. Buckmaster. 2006. *Engineering Principles of Agricultural Machinery.* St. Joseph, MI: American Society of Agricultural and Biological Engineers.

Tong, X. and Z. Wang. 2012. Fuzzy Acceptance Sampling Plans for Inspection of Geospatial Data with Ambiguity in Quality Characteristics. *Computers & Geosciences* 48256-266.

Tong, X., Z. Wang, H. Xie, D. Liang, Z. Jiang, J. Li and J. Li. 2011. Designing a Two-Rank Acceptance Sampling Plan for Quality Inspection of Geospatial Data Products. *Computers & Geosciences* 37(10): 1570-1583.

United States Environmental Protection Agency. 2003. Guidance for Geospatial Data Quality Assurance Project Plans. United States Environmental Protection Agency. Available at: http://www.epa.gov/quality/qs-docs/g5g-final.pdf. Accessed April 17 2013.

Yahya, A., M. Zohadie, A. F. Kheiralla, S. K. Giew and N. E. Boon. 2009. Mapping System for Tractor-Implement Performance. *Computers and Electronics in Agriculture* 69(1): 2-11.

Yan, G., J. Wang and S. Chen. 2011. Performance Analysis for (X,S)-Bottleneck Cell in Large-Scale Wireless Networks. *Information Processing Letters* 111(6): 269-277.

Yule, I. J., G. Kohnen and M. Nowak. 1999. A Tractor Performance Monitor with DGPS Capability. *Computers and Electronics in Agriculture* 23(2): 155-174.

APPENDIX A: .csv FILE MERGING PROGRAM

The following code was written by Andy Stevens with John Deere in order to merge the individual files for each instance of the machine being turned on into one master file for each machine.

```
# Merge all MTG files in a directory
# Change path name in 'setwd' (= set Working Directory)
# make sure only MTG data csv files are located there (i.e. no "P83605680B.csv" left over)
setwd("C:\ […] \csv_files")
list.files(pattern = "*.csv")   # check that you have the right directory and there are csv files there
OutFileName <- "P61701260C.csv" # output you want file name - update on each run
# library(plyr) # required for sorting dataframe to make sure in the correct order after complex merging
read.MTG.file <- function(filename) {
  # read CAN data in csv file format
  # extract the ID, Date and Time from the filename
  # example:  "K83455174_20120424_132920.csv"
  fileNameLen    <- nchar(filename)    # file Name Length
  fileTimeEnd    <- fileNameLen    - 4 # file Time End position
  fileTimeStart  <- fileTimeEnd    - 5 # file Time Start position
  fileDateEnd    <- fileTimeStart  - 2 # file Date End position
  fileDateStart  <- fileDateEnd    - 7 # file Date Start position
  fileIDEnd      <- fileDateStart  - 2 # file ID End position
  fileIDStart    <- 1                  # file ID Start position
  fileID   <- substring(filename, fileIDStart  , fileIDEnd  )
  fileDate <- substring(filename, fileDateStart, fileDateEnd)
  fileTime <- substring(filename, fileTimeStart, fileTimeEnd)
  # read the header on line 1
```

```
  filehead <- read.csv(filename, header=F, nrows=1, as.is = TRUE)
  # read in the data starting on line 3
  filedata <- read.csv(filename, header=F, skip=2)
  # assign variable names and file ID, Date and Time
  filehead <- make.unique(as.character(filehead)) # make sure variable names are unique
  names(filedata) <- filehead
  filedata$fileID <- fileID
  filedata$fileDate <- fileDate
  filedata$fileTime <- fileTime
  filedata
}
read.MTG <- function (combine.method = "rbind")
{
  # read MTG csv files in a directory
  # use combine.method = "rbind" if files have the same variables (the default, hopefully)
  # use combine.method = "merge" if discover files have different variables (might take
longer and have to sort later or add to this script)
  filelist <- list.files(pattern = "*.csv")
  for (i in 1:length(filelist)) {
    filedata = read.MTG.file(filelist[i])
    if (i == 1) {
      batchdata = filedata
    }
    else if(combine.method == "rbind") {
      batchdata = rbind(batchdata, filedata)
    }
    else {
      batchdata = merge(batchdata, filedata, all = TRUE)
    }
  }
```

```
  batchdata
}
MTGfile <- read.MTG("rbind") # try with "rbind" and if it crashes, change to "merge"
names(MTGfile) # just to check the variable names
colnames(MTGfile) <- gsub(" ", "_", colnames(MTGfile)) # take out any spaces in the variable names
# fix misspelling in GPS_Altitude
# MTGfile <- within(MTGfile, {
#  GPS_Altitude <- GPS_Altitiude
#  rm(GPS_Altitiude)
#  }
#)
# MTGfile <- arrange(MTGfile, fileDate, fileTime, Sample) # just to make sure they are in the right order
write.csv(MTGfile, OutFileName)
rm(list = ls()) # optional - remove objects to save space
```

APPENDIX B: DATA LOSS BY MTG FILE SETUP PER MACHINE

Table 8 – Data loss associated with Farm 1 by data collection device file setup with a summary of all machines (expansion of Table 1)

Farm 1									
Machine	File Setup	File Size (kB)		Data Loss (%)	Machine	File Setup	File Size (kB)		Data Loss (%)
		Raw	Cleaned				Raw	Cleaned	
JD 4940	a	70,106	34,367	51%	JD 8360-5	a	18,799	10,135	46%
JD 4940	b	1,706	1,341	21%	JD 8360-5	b	103,428	58,177	44%
JD 6170-1	a	31,397	18,552	41%	JD 8360-5	c	34,679	5,863	83%
JD 6170-1	b	168,594	74,537	56%	JD 8360-5	d	241,011	73,277	70%
JD 6170-2	a	36,979	18,604	50%	JD 9460-1	a	37,381	18,415	51%
JD 6170-2	b	110,189	50,286	54%	JD 9460-1	b	157,076	84,210	46%
JD 8360-1	a	5,527	1,883	66%	JD 9460-1	c	231,253	157,822	32%
JD 8360-1	b	108,531	67,077	38%	JD 9460-2	a	9,067	5,463	40%
JD 8360-1	c	74,248	52,988	29%	JD 9460-2	b	141,605	75,751	47%
JD 8360-1	d	2,023	1,214	40%	JD 9460-2	c	336,563	190,932	43%
JD 8360-1	e	57,163	48,952	14%	JD 9460-3	a	27,167	16,411	40%
JD 8360-2	a	26,619	15,592	41%	JD 9460-3	b	117,016	56,004	52%
JD 8360-2	b	157,276	98,056	38%	JD 9460-3	c	88,707	25,547	71%
JD 8360-2	c	38,664	23,740	39%	JD 9460-4	a	24,839	12,126	51%
JD 8360-2	d	46,121	25,056	46%	JD 9460-4	b	142,572	74,989	47%
JD 8360-2	e	172,275	110,504	36%	JD 9460-4	c	10,201	7,001	31%
JD 8360-3	a	9,732	5,225	46%	JD 9460-5	a	19,838	9,222	54%
JD 8360-3	b	108,226	61,524	43%	JD 9460-5	b	172,336	117,488	32%
JD 8360-3	c	93,686	57,069	39%	JD 9870	a	51,572	36,708	29%
JD 8360-3	d	49,919	37,979	24%	JD S680-1	a	238,469	169,503	29%
JD 8360-4	a	41,591	26,104	37%	JD S680-2	a	24,820	16,762	32%
JD 8360-4	b	69,714	33,847	51%	JD S680-2	b	242,812	125,979	48%
JD 8360-4	c	102,368	63,528	38%					
JD 8360-4	d	268,808	137,394	49%	***All Machines***		***4,322,673***	***2,413,204***	***44%***

Table 9 – Data loss associated with Farm 2 by data collection device file setup with a summary of all machines (expansion of Table 2)

Farm 2				
Machine	File Setup Version	File Size (kB)		Data Loss (%)
		Raw	Cleaned	
JD 4830	a	43,943	22,314	49%
JD 4830	b	291,891	190,737	35%
JD 4830	c	39,924	26,072	35%
JD 8345-1	a	62,046	39,197	37%
JD 8345-1	b	52,788	28,266	46%
JD 8345-1	c	241,713	126,214	48%
JD 8345-1	d	217,172	187,570	14%
JD 8345-2	a	61,010	39,109	36%
JD 8345-2	b	240,528	139,148	42%
JD 8345-3	a	163,413	97,840	40%
JD 8345-3	b	175,541	151,456	14%
JD 9770-1	a	471	312	34%
JD 9770-1	b	69,035	46,902	32%
JD 9770-2	a	71,402	48,298	32%
JD 9770-2	b	124,248	25,304	80%
JD 9770-3	a	66,083	45,156	32%
JD 9770-3	b	249,263	203,278	18%

All Machines	***2,170,471***	***1,417,173***	***35%***

APPENDIX C: ORDINARY LEAST SQUARES OUTPUT EXAMPLES

C.1 Output Files from ArcGIS ™ for Multiple Variable Ordinary Least Squares (OLS)

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	-27.006236	0.394175	-68.513310	0.000000*	0.516162	-52.321231	0.000000*	--------
CAN_ENGINE	0.029304	0.000221	132.694653	0.000000*	0.000326	89.834941	0.000000*	1.800596
GROUNDSPEE	0.517997	0.025397	20.396268	0.000000*	0.039312	13.176516	0.000000*	1.801755
PITCH	2.600890	0.038868	66.916012	0.000000*	0.045374	57.321513	0.000000*	1.003081

Figure 24 – ArcGIS™ Ordinary Least Squares results summary (multiple variable model)

OLS Diagnostics

Input Features:	K97707456B.shp	Dependent Variable:	CAN_FUELRA
Number of Observations:	19163	Akaike's Information Criterion (AICc) [d]:	123975.365588
Multiple R-Squared [d]:	0.700263	Adjusted R-Squared [d]:	0.700216
Joint F-Statistic [e]:	14920.124763	Prob(>F), (3,19159) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	36562.785643	Prob(>chi-squared), (3) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	2.886063	Prob(>chi-squared), (3) degrees of freedom:	0.409528
Jarque-Bera Statistic [g]:	7459.450279	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Figure 25 – ArcGIS™ Ordinary Least Squares diagnostics (multiple variable model)

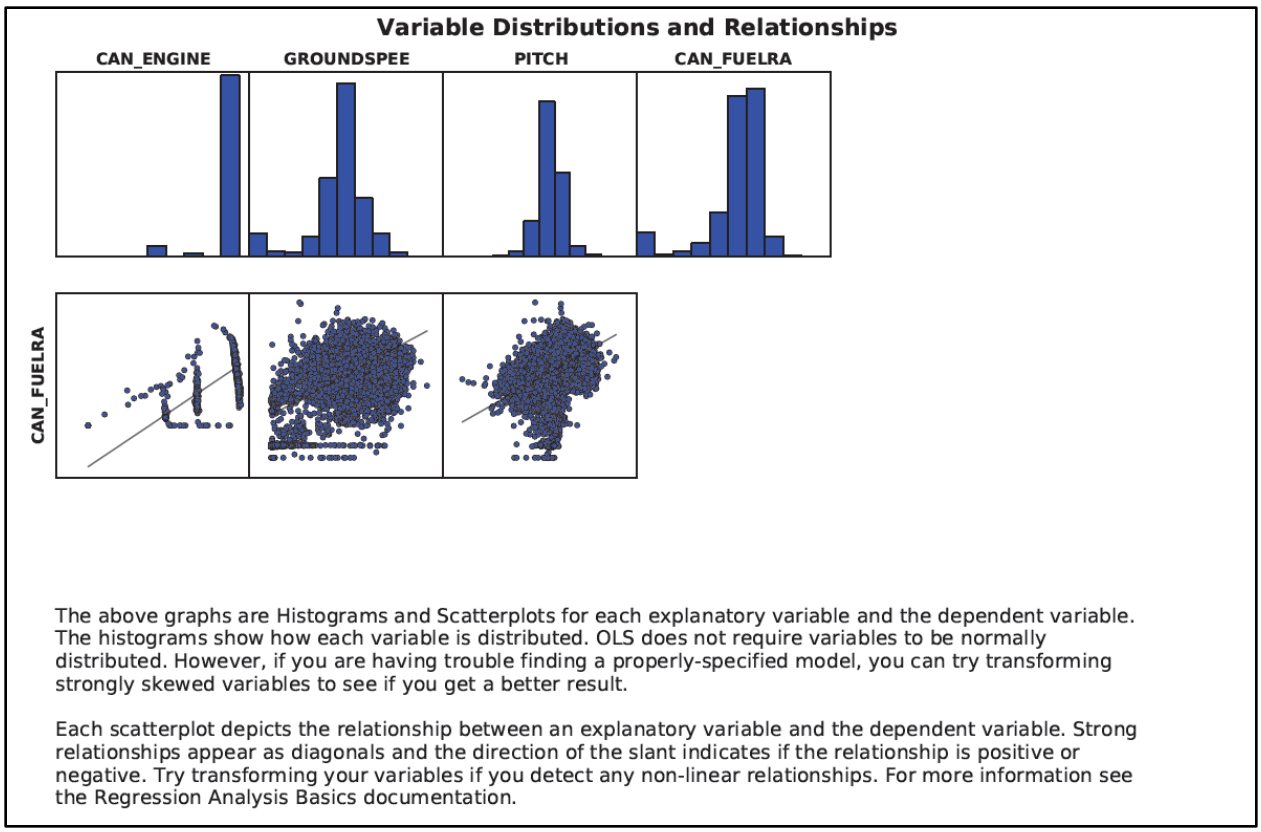


Figure 26 – ArcGIS™ Ordinary Least Squares variable distributions and relationships (multiple variable model)

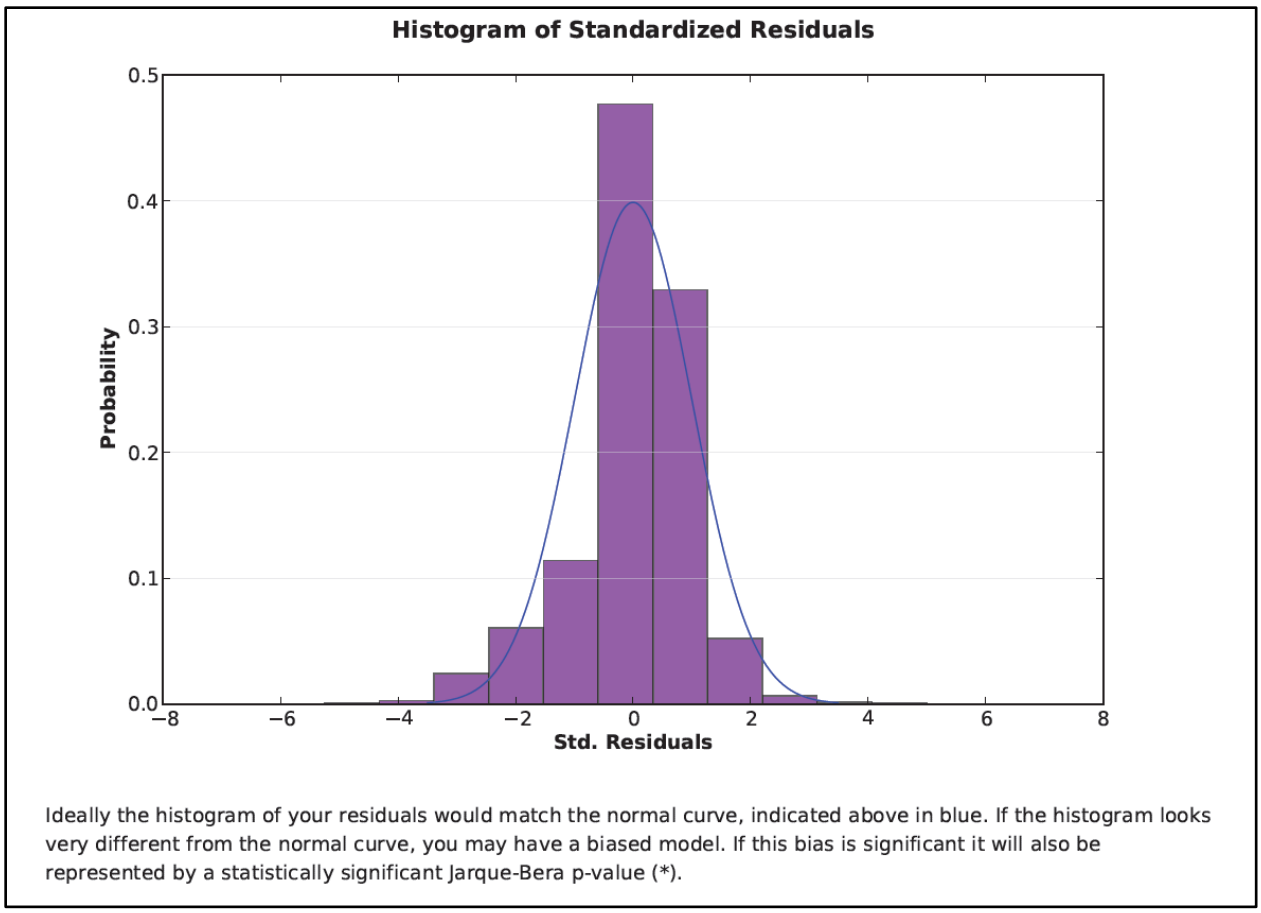


Figure 27 – ArcGIS™ Ordinary Least Squares histogram of standardized residuals (multiple variable model)

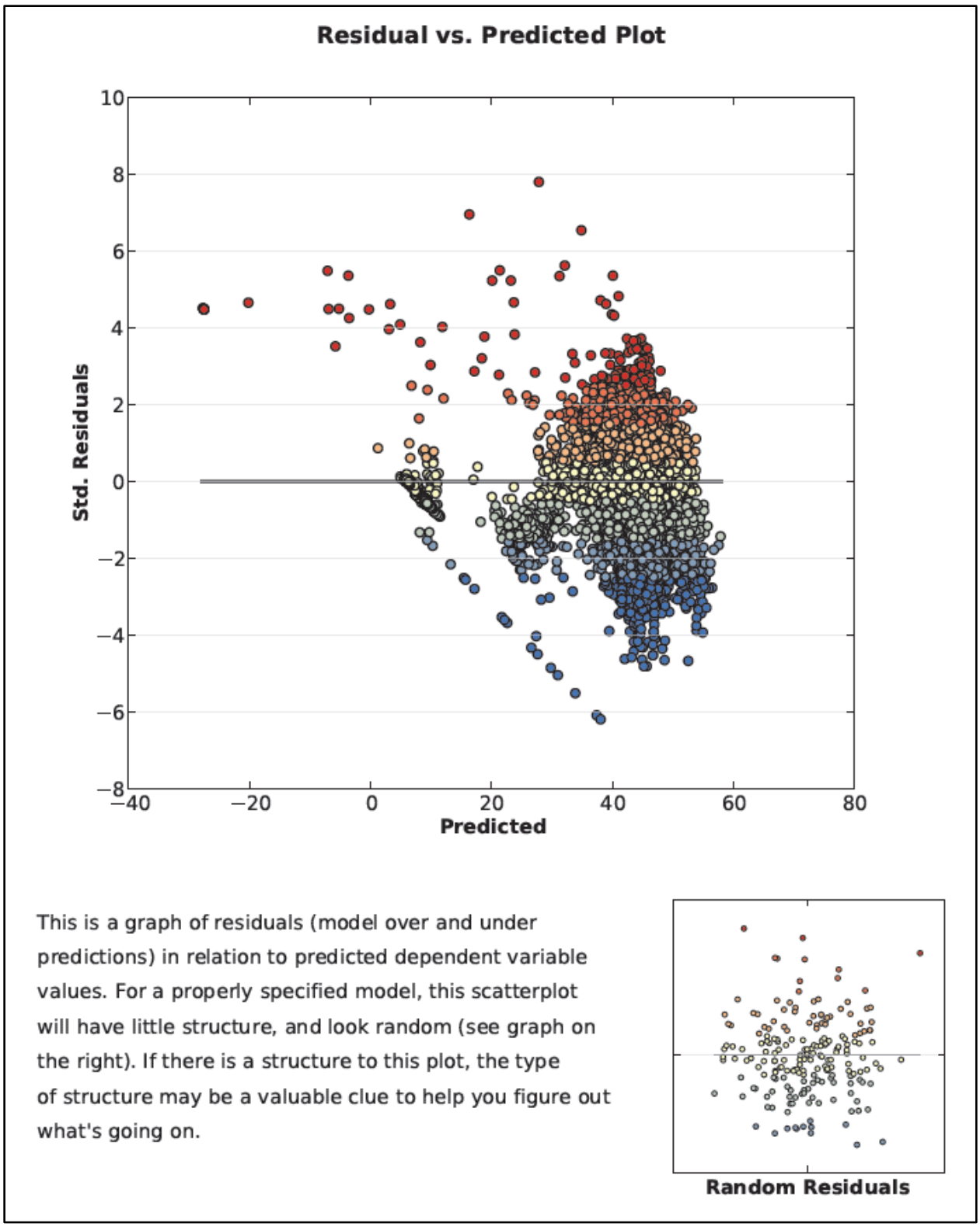


Figure 28 – ArcGIS™ Ordinary Least Squares residual vs. predicted plot (multiple variable model)

C.2 Output Files from ArcGIS ™ for the Single Variable Ordinary Least Squares (OLS)

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Intercept	-31.249901	0.413218	-75.625657	0.000000*	0.451397	-69.229234	0.000000*
CAN_ENGINE	0.032826	0.000185	177.705043	0.000000*	0.000203	161.499931	0.000000*

Figure 29 – ArcGIS™ Ordinary Least Squares results summary (single variable model)

OLS Diagnostics

Input Features:	K97707456B.shp	Dependent Variable:	CAN_FUELRA
Number of Observations:	19163	Akaike's Information Criterion (AICc) [d]:	128398.222119
Multiple R-Squared [d]:	0.622370	Adjusted R-Squared [d]:	0.622350
Joint F-Statistic [e]:	31579.082367	Prob(>F), (1,19161) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	26082.227689	Prob(>chi-squared), (1) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	18.217126	Prob(>chi-squared), (1) degrees of freedom:	0.000020*
Jarque-Bera Statistic [g]:	3122.313175	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.05$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.05$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.05$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.05$) model predictions are biased (the residuals are not normally distributed).

Figure 30 – ArcGIS™ Ordinary Least Squares diagnostics (single variable model)

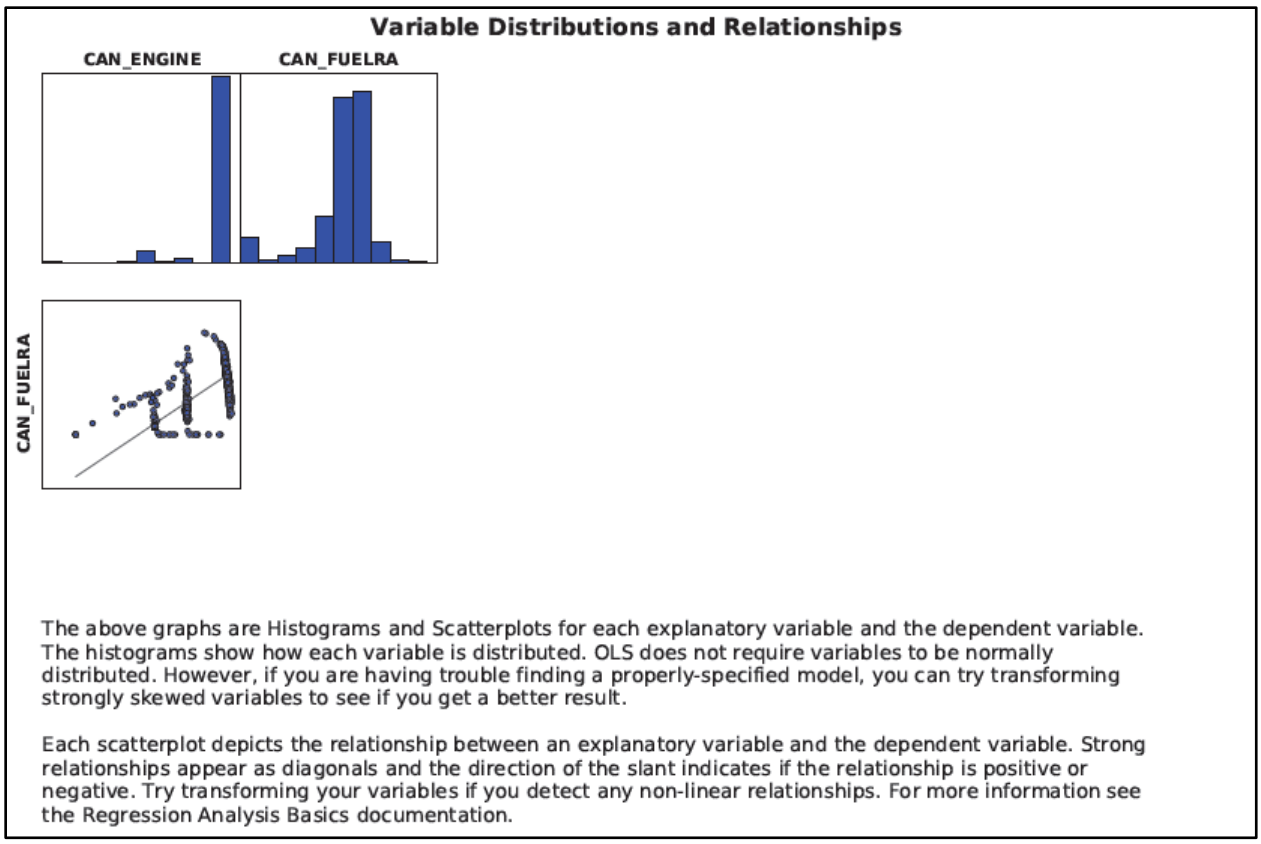


Figure 31 – ArcGIS™ Ordinary Least Squares variable distributions and relationships (single variable model)

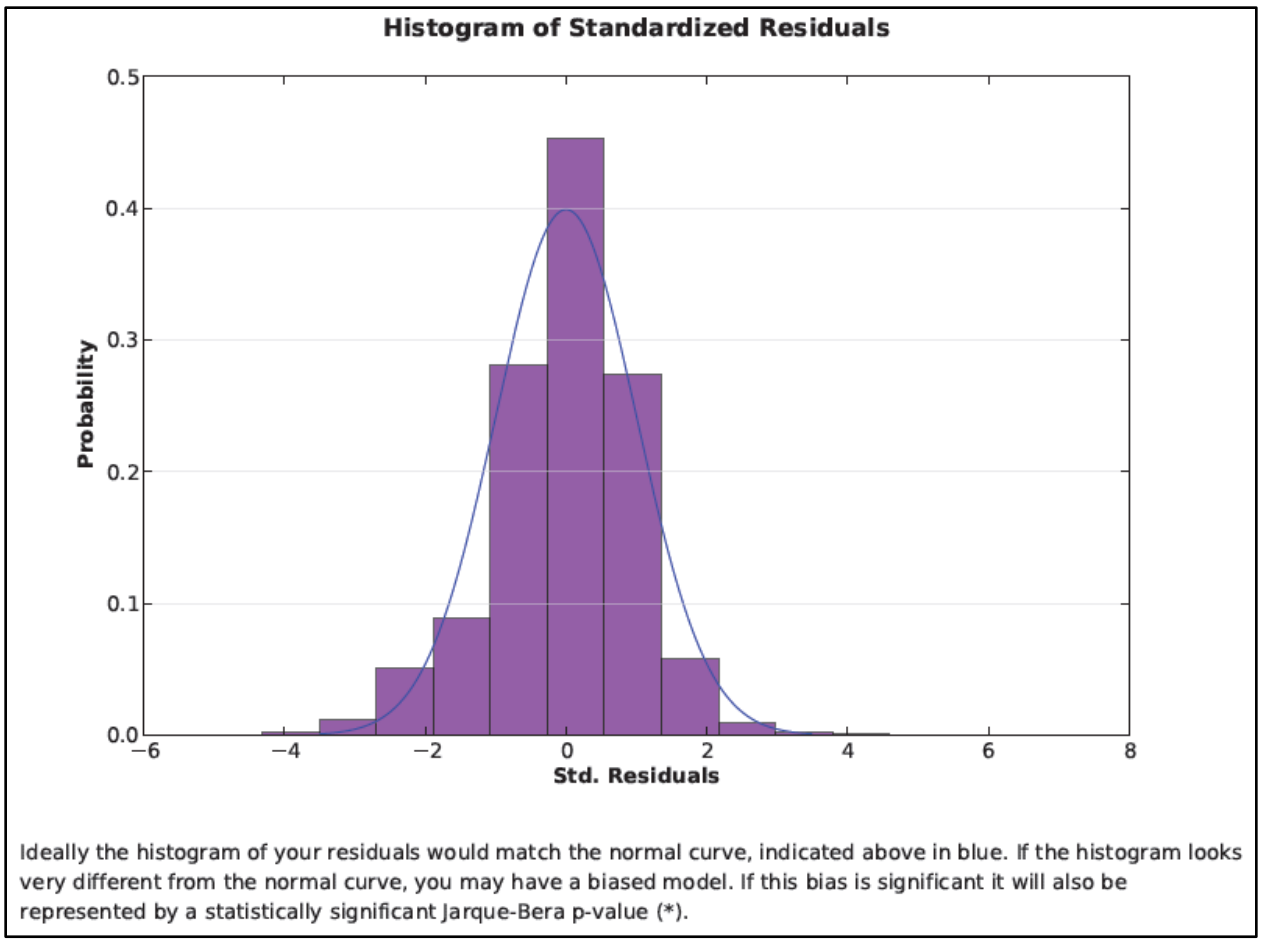


Figure 32 – ArcGIS™ Ordinary Least Squares histogram of standardized residuals (single variable model)

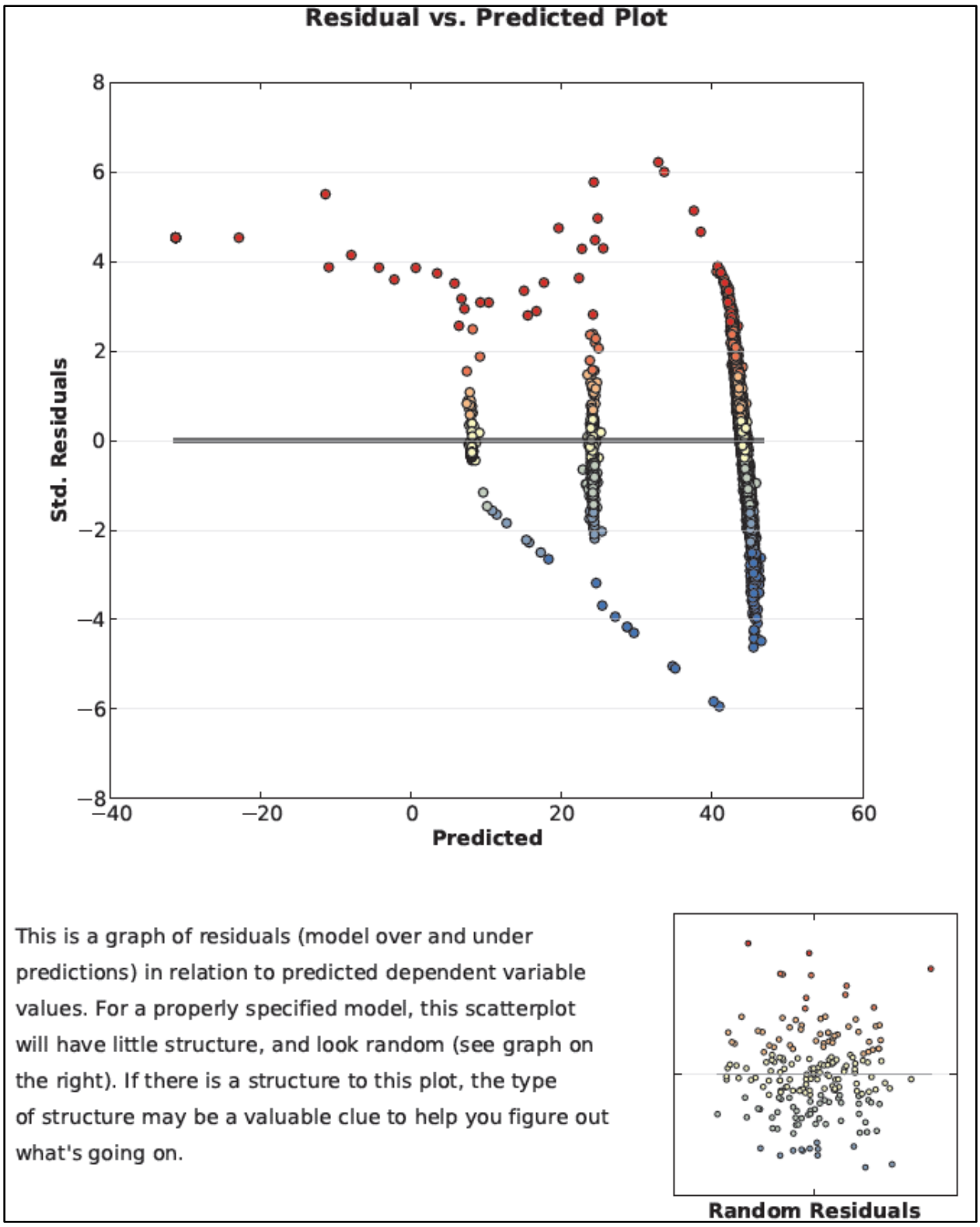


Figure 33 – ArcGIS™ Ordinary Least Squares histogram of residual vs. predicted plot (single variable model)

APPENDIX D: REVISED MACHINE STATE RULES

The following descriptions of machine state created by Niehaus (2013) are provided as recommended improvements to the definitions of machine states presented in this study. If these parameters can be collected accurately, the application of these rules will provide a more accurate assessment of machine state (Niehaus 2013).

Table 10 – Recommended definitions for machine state during tillage operations if additional parameters are collected in the future

Tillage States				
Lat/Lon (Degrees)	Engine Speed (rpm)	Ground Speed, kph (mph)	Implement Position	State
Farmstead	0-1500	0-3.2 (0-2)	Up	Idle
In-Transit	1500-2200	13.7+ (8.5+)	Up	Transport
In-Field 1	1500-2200	3.2-13.7 (2-8.5)	Down	Tilling
In-Field 1	1500-2200	0-3.2 (0-2)	Down	Abnormal

Table 11 – Recommended definitions for machine state during planting operations if additional parameters are collected in the future

Planting States				
Lat/Lon (Degrees)	Engine Speed (rpm)	Ground Speed, kph (mph)	Planter Position	State
Farmstead	0-1500	0-3.2 (0-2)	Up	Idle
In-Transit	1300-2200	12.1+ (7.5+)	Up	Transport
In-Field 1	1300-2200	6.4-8 (4-5)	Down	Tilling
In-Field 1	1300-2200	Speed < 6.4 (4) or >8 (5)	Down	Abnormal

APPENDIX D: REVISED MACHINE STATE RULES (Cont.)

Table 12 – Recommended definitions for machine state during harvest operations if additional parameters are collected in the future

Harvest States							
Lat/Lon (Degrees)	Heading (Degrees)	Engine Speed (rpm)	Ground Speed, kph (mph)	Head Position	Separator	Unloading Auger	State
Farmstead	N/A	0-1500	0-3.2 (0-2)	Up	Off	Off	Idle
In-Transit	Variable	1500-2200	10.5 (6.5+)	Up	Off	Off	Transport
In-Field 1	N/A	0-1500	0-3.2 (0-2)	Up	Off	Off	Idle
In-Field 1	Constant	1500-2200	3.2-10.5 (2-6.5)	Down	On	Off	Harvesting
Edge-Field 1	Variable	1500-2200	3.2-10.5 (2-6.5)	Up	On	Off	Turning
In-Field 1	Constant	1500-2200	3.2-10.5 (2-6.5)	Down	On	On	Harvesting and Unloading
Edge-Field 1	Constant	1500-2200	0 (0)	Down	On	On	Stopped and Unloading
In-Field 1	Variable	1500-2200	10.5+ (6.5+)	Up	Off	Off	Transport
In-Field 1	N/A	1500-2200	0-3.2 (0-2)	Down	On	Off	Abnormal